7-2014

Functional Genomics of Maize Endosperm Maturation and Protein Quality

Lingling Yuan
University of Nebraska-Lincoln

FUNCTIONAL GENOMICS OF MAIZE ENDOSPERM MATURATION AND PROTEIN QUALITY

by

Lingling Yuan

A DISSERTATION

Presented to the Faculty of

The Graduate College at the University of Nebraska

In Partial Fulfillment of Requirements

For the Degree of Doctor of Philosophy

Major: Agronomy and Horticulture

(Plant Breeding and Genetics)

Under the Supervision of Professor David R Holding

Lincoln, Nebraska

July, 2014

FUNCTIONAL GENOMICS OF MAIZE ENDOSPERM MATURATION AND PROTEIN QUALITY

Lingling Yuan, Ph.D.

University of Nebraska, 2014

Advisor: David R Holding

Maize is one of the most important cereal crops and widely cultivated throughout the world. The study on maize kernel development including protein quality improvement is essential for removing dietary protein deficiency because of the lack of essential amino acids, especially lysine and tryptophan, in maize kernel. Quality Protein Maize (QPM) is a hard kernel variant of the high-lysine mutant, *opaque-2*. We created opaque QPM variants to identify *opaque-2* modifier genes and to investigate deletion mutagenesis combined with Illumina sequencing as a maize functional genomics tool. A K0326Y-QPM deletion mutant, line 107, was null for the 27- and 50-kD γ-zeins and abolished vitreous endosperm formation. Characterization of line 107 identified 27-kD γ-zein as an *opaque-2* modifier gene within the largest QPM QTL, and may suggest the 50-kD γ-zein also contributes to this QTL. It further demonstrates that genome-wide deletions in non-reference maize lines can be identified through a combination of assembly of Illumina reads against the B73 genome and integration of RNA-seq data. Characterization of a low α-zein mutant, line 198, mapped its mutation on chromosome 3.02 and I describe our efforts towards its molecular identification.

Besides the K0326Y-QPM mutagenesis project, I made a significant contribution to the

study of zein function and their inverse relationship to kernel protein quality by using RNA interference. I addressed the extent of functional non-redundancy within the γ- and α-zein subclasses. My involvement in this work was in-depth analysis of transcript and protein levels in endosperms of these lines.

I was also involved in the characterization of a sorghum *high digestibility high lysine* opaque mutant *hdhl.* We found that the mutant results from the missense mutation results in the 21st conserved amino acid alanine (Ala) in the signal peptide being replaced with threonine (Thr), which renders the signal peptide resistant to processing, indirectly reducing the levels of lysine-poor kafirins and thereby increasing lysine-rich proteins and protein digestibility in the seeds.

Acknowledgements

First of all, I would like to express my sincerest appreciation to my advisor Dr. David Holding. His great mentorship, support and constructive instructions help me through my whole Ph.D. program. His distinct scientific expertise and deep devotion in science encouraged me so much and showed how to become a real scientist. My advisor definitely is my career example. Great thanks to all of my committee members, Dr. Staswick, Dr. Waters, and Dr. Zhang who have provided invaluable advice and encouragement for improvements of my research during my study in UNL. I appreciate all your patience and help.

I also deeply appreciate all of our lab members. I have learned so much from our former Post doctor Xiaomei Guo and our present member Kyla Morton, which I could never learn from the books. And they help me a lot in fieldwork during each sultry tropical heat summer. Big thanks to all of the friends I have made in the Beadle center, especially Shuxin Zhang, Yongchao Dou, Guodong Ren, Fan Yang, Xiangjun Li. I really appreciate our precious friendship and all the suggestions you provide both in research and life.

Finally, I am eternally grateful to my family members. Thanks to my father and mother for loving me unconditionally. Thanks to my husband Jianxin Sun for his great help and care. I dedicate this dissertation to all of my family members for their endless love and support.

Funding Acknowledgement

This research was mainly supported by the UNL Department of Agronomy and Horticulture and Center for Plant Science Innovation, and during the last phase, by the Agriculture and Food Research Initiative competitive grant no. 2013-02278 of the USDA National Institute of Food and Agriculture

Table of Contents

TITLE PAGE……i

ABSTRACT……ii

COPYRIGHT……iv

ACKNOWLEDGEMENTS……v

GRANT INFORMATION……..……...vi

TABLE OF CONTENTS…….…….……vii

LISTS OF MULTIMEDIA OBJECTS

CHAPTER 1……...ix

CHAPTER 2……...……ix

CHAPTER 3……...……...…..x

APPENDIX 1……. xi

APPENDIX 2……....xii

CHAPTER 1 Introduction

Introduction…….2

Figures……23

References…….....25

CHAPTER 2 Deletion mutagenesis identifies a haploinsufficient role for gamma-zein in *opaque 2* endosperm modification

Abstract…….....40

Introduction……..……..40

Methods……..43

Results…….....49

Discussion……...……...58

Figures……65

References……..72

CHAPTER 3 Characterization of K0326Y deletion line 198: a QPM opaque revertant with improved protein quality

Abstract……79

Methods……80

Results……...81

Discussion……...…..85

Figures…….....88

References……94

APPENDIX 1……...96

APPENDIX 2……..105

Lists of Multimedia Objects

CHAPTER 1 Introduction

Figure 1. Phases in maize endosperm development.

Figure 2. Current model for zein protein body architecture at early, mid- and mature stages.

Table 1. Opaque mutants summary

CHAPTER 2 Deletion mutagenesis identifies a haploinsufficient role for gamma-zein in *opaque 2* endosperm modification

Figure 1. M3 ear phenotypes of K0326Y QPM control ear and segregating and homozygous K0326Y deletion line 107. Inserts in A and C show light box phenotypes.

Figure 2. zein profiles of mutants compared with W64A wild type, *W64A o2* and K0326Y QPM.

Figure 3. A, SDS-PAGE analysis of zein proteins from mature kernel of deletion line 107 compared with K0326Y QPM and the W64A wild type and *W64Ao2*. B, Western analysis of zein proteins loaded at 1:1,000 dilution of gel shown in A, probed with total α-zein antiserum (1:10,000) and anti-27-kD γ-zein antiserum (1:2,000). WT, Wild type.

Figure 4. Low-magnification TEM images from immunogold labeling showing distribution of protein bodies in representative 18-DAP endosperm of the wild type, K0326Y QPM, and QPM deletion line 107.

Figure 5. Immunogold TEM analysis showing α- and γ-zein distribution in protein bodies of fourth sub-aleurone cell layer of 18-DAP endosperm of the wild type, K0326Y QPM, and QPM deletion line 107.

Figure 6. A, RT-PCR showing γ-zein transcript accumulation in line 107.
B, Genomic PCR analysis of γ-zein genes in line 107.

Figure 7. SDS-PAGE and kernel opacity analysis showing haploinsufficiency of γ-zein for endosperm modification. Selected kernels from a segregating ear are shown.

Figure 8. SDS-PAGE showing zein accumulation and its relationship to kernel vitreousness in F3 kernels resulting from a line 107xB73 wild type out-cross

Figure 9. Exon-seq and genomic and RT-PCR verification of deletion on chromosome 7.02.

Figure 10. Exon-seq and genomic and RT-PCR verification of deletion on chromosome 9.03.

Table 1. Protein body number in fourth sub-aleurone layer of the wild type, K0326Y QPM, and K0326Y line 107.

Table II. Lys and Trp concentrations in mature kernels.

CHAPTER 3 Characterization of K0326Y deletion line 198: a QPM opaque revertant with improved protein quality

Figure 1. M3 ear phenotypes of K0326Y QPM control ear and segregating and homozygous K0326Y deletion line 198. Inserts in A and C show light box phenotypes.

Figure 2. A, SDS-PAGE analysis of zein proteins from mature kernel of deletion line 198 compared with K0326Y QPM and the W64A wild type and *W64Ao2*. B, Western analysis of zein proteins loaded at 1:1,000 dilution of gel shown in A, probed with total α-zein antiserum (1:10,000) and 27-kD γ-zein antiserum (1:2,000).

Figure 3. A, Genomic PCR of Z1 locus in both QPM non-mutagenesis line and 198. B, RT-PCR showing Z1 gene transcript accumulation in 198 compared with QPM non-mutagenesis. Primers are sub-family specific and work for both gPCR and RT-PCR.

Figure 4. Expression analysis of zein genes by quantitative real time RT-PCR in 198 relative to QPM non-mutagenesis. Each graph represents the measurement of a different zein gene in this mutant. Primers are sub-family specific (Feng *et al.,* 2009).

Figure 5. Southern Blot analysis of α-zeins genes in 5 line 198 mutants compared with 15 controls. Probe: mix of z1A B C D zein gene family-specific primers.

Figure 6. Bulked Segregant analysis mapping of mutation in mutant line 198.

Figure 7. Genomic distribution of α-zeins genes in B73 and BSSS53 inbred lines.

Table 1. Lys and Trp concentrations in mature kernels of 198.

APPENDIX 1

Figure 1. Diagram of zein RNAi constructs with 27-kD γ-zein promoter (p27γ) and Cauliflower mosaic virus terminator (T35S).

Figure 2. qRT-PCR expression analysis of α- and γ-zein genes in RNAi lines relative

to wild-type non-transgenic (HiII).

Figure 3. TEM analysis showing protein body density in fourth sub-aleurone starchy layer of 18-DAP endosperm in zein RNAi lines and their crosses.

Figure 4. TEM analysis showing protein size and morphology in fourth sub-aleurone starchy layer of 18-DAP endosperm in zein RNAi lines and their crosses.

APPENDIX 2

Figure 1. Alignment of the signal peptides of α-prolamins from the grass subfamily.

Figure 2. Missense mutation at the 21st amino acid of the signal peptide in maize *floury2* mutants

Figure 3. Missense mutation at the 21st amino acid of the signal peptide in *hdhl*.

Figure 4. Western blot analysis of kafirin accumulation in maize null segregants, mutant kafirin transgenic seeds and the transgenic seeds expressing normal 22-kD α-kafirins.

CHAPTER 1.

INTRODUCTION

1 The importance of maize

Maize, the domesticated variant of teosinte, has a single tall stalk with multiple leaves while teosinte is a short bushy plant. Maize is one of the most important cereal crops and widely cultivated throughout the world. Maize, rice and wheat constitute more than 80% of total global cereals production (Sofi et al., 2009). In industry, maize is increasingly used for the production of ethanol, an essential biofuel (Li et al., 2014). The increasing use of bioethanol has been pushed for years as a replacement for fossil fuels. More than 90% of bioethanol in the world is produced from maize, of which the majority is derived from starch from the grain. In addition, the stalk, leaves, cob and husk parts of maize plant represent 50 to 70% of its total dry matter. These non-grain parts can be used for silage production and serve as a valuable feed source for ruminant animals (Hunt et al., 1989). Most importantly, besides the important role of the grain portion and other plant parts on industrial and silage production respectively, maize is a staple food for many millions of people and their livestock. It often becomes the main food source when other food sources are limited in developing and underdeveloped nations. However, such dependence on maize puts people and monogastric animals at risk for dietary protein deficiency because of the lack of essential amino acids, especially lysine and tryptophan, in the maize kernel. Thus, the study of maize kernel development, including protein quality improvement, is of great importance.

2 Endosperm development in maize

Endosperm development in maize, as shown in Figure 1, includes several distinct phases: double fertilization, syncytium formation, cellularization, differentiation, accumulation of

storage compounds and maturation (programmed cell death, dormancy and desiccation) (Kranz et al., 1998; Sabelli and Larkins, 2009; Sabelli et al., 2013). Double fertilization creates endosperm and embryo. The embryo derives from the fertilization of one sperm nucleus with the egg cell nucleus and the endosperm results from the fusion between a sperm and two polar nuclei (Kiesselbach, 1949; Kiesselbach, 1999). The triploid nucleus undergoes several rounds of synchronous division to form a syncytium without the formation of cytoplasm, cell membranes, cell wall and cytokinesis (Sabelli and Larkins, 2009). When there already are four to eight endosperm nuclei, the zygote divides for the first time. This suggests greater mitotic activity in the triploid primary endosperm nucleus than the cell division in the zygote in maize (Randolph, 1936). The coenocytic endosperm undergoes cellularization from the periphery of the endosperm toward the central vacuole. The nuclear-cytoplasmic domains are defined by the formation of microtubule systems that radiate from the nuclear surface (Olsen et al., 1999). Cell wall formation is initiated at sites of microtubule intersection and deposition of an adventitious phragmoplast. The centripetal extension of the cell walls results in alveoli. The nucleus divides synchronously and periclinally, which is followed by cytokinesis (Sabelli and Larkins, 2009).

Four major cell types including transfer cells, aleurone cells, endosperm cells, and embryo-surrounding region cells constitute the mature maize endosperm (Olsen, 2001; Sabelli and Larkins, 2009). Generally, the transfer cell layers are located over the chalazal pad and stop dividing and differentiate early. They facilitate endosperm nutrient uptake by their extensive cell wall invaginations and increased plasma membrane surface

area. The differentiation of the transfer cells requires high metabolic rates and typically, they have a dense cytoplasm that is rich in mitochondria (Charlton et al., 1995; Sabelli and Larkins, 2009). The outer region of the maize endosperm consists of a single layer of aleurone cells that differentiates between six and 10 days after pollination (DAP) from the outer layer and surrounds the endosperm except in the transfer cell region (Sabelli and Larkins, 2009; Becraft and Yi, 2010). The fate of aleurone cells is specified soon after the first periclinal division of the cellularized endosperm. Aleurone cells contain numerous small vacuoles with aleurone grains and anthocyanin. Aleurone grains consist of bounding membrane enclosing a matrix in which are embedded spherical deposits (Buttrose, 1963). And anthocyanin imparts the mature maize kernel colors (Olsen, 2001; Sabelli and Larkins, 2009). At endosperm maturity, the aleurone is the only living tissue and not desiccated through a specific mechanism. During imbibition, aleurone cells synthesize proteolytic and hydrolytic enzymes that can digest the endosperm cell walls, which causes the starch and proteins stored in the endosperm cells to be recycled by the embryo (Gibbon and Larkins, 2005; Sabelli and Larkins, 2009).

Maize endosperm, containing around 90% starch and 10% protein in terms of dry weight，is one of the most important sources of food calories and protein for human beings and livestock feed worldwide (Gibbon and Larkins, 2005; Li et al., 2014). Starch, including amylose and amylopectin, is packed into granules within amyloplasts. Four enzymes function in starch biosynthesis. Three of them, including starch synthase, starch-branching and starch-debranching enzyme, are found in amyloplasts (Smith, 1999; James et al., 2003). Another enzyme called ADP-Glc pyrophosphorylase controls the rate-

limiting step in starch synthesis and is mainly found in the cytosol of the endosperm cells and facilitates starch biosynthesis in the presence of sucrose. Starch grains start to accumulate in the maize endosperm at a round 10 DAP (Smith, 1999; James et al., 2003; Hannah, 2007; Sabelli and Larkins, 2009). Protein, the other main storage component in maize endosperm, accumulates as a reservoir of amino acids for the embryo during pre-photosynthetic growth. Endosperm protein contributes to the biophysical properties of the mature endosperm tissue and is important in cereal end-use properties such as kernel vitreousness in the case of maize (Holding, 2014) and viscoelastic properties of bread dough in the case of wheat (Shewry and Tatham, 1990; Sabelli and Larkins, 2009). The principle storage proteins in cereals, prolamins, are highly hydrophobic and soluble in 70% alcoholic solutions or denaturing solvents. Maize prolamins are called zein and represent more than 70% of total endosperm proteins. Zeins form insoluble accretions called protein bodies in the lumen of the rough endoplasmic reticulum (RER) where they are retained (Coleman and Larkins, 1999; Holding and Larkins, 2006; Holding and Messing, 2013). The organization of prolamins in protein bodies is directed in part by regulation of zein gene expression and specific interactions between each zein fraction (Woo et al., 2001; Kim et al., 2002). Specific mutations in prolamins disrupt the organization of protein bodies and lead to the unfolded protein response (UPR) (Holding, 2014). UPR is a cellular stress response that is activated in response to an accumulation of unfolded or misfolded proteins in the lumen of the endoplasmic reticulum. Though zeins account for more than 70% of maize endosperm protein, they are devoid of the essential amino acids lysine and tryptophan, which causes the overall protein deficiency of these amino acids in the grain (Holding and Larkins, 2006; Holding and Messing, 2013).

3 Zeins and protein body formation

The maize starchy endosperm usually consists of a vitreous (glassy) region at the periphery of the mature kernel and a central opaque region (Gibbon and Larkins, 2005). Vitreous endosperm is important for resistance to insect and fungal damage and resilience during harvest and storage. The accumulation and packaging of zeins into ER protein bodies play a central role in the formation of the vitreous endosperm. The vitreous outer region of the maize kernel contains much more zein than the opaque central region. Nitrogen deficiency in the plant causes reduced zein synthesis and loss of the vitreous region (Tsai et al., 1978). Protein body formation is controlled by the regulation of zein gene expression at both the temporal and spatial levels, and by interactions between the different zein components (Woo et al., 2001; Guo et al., 2013).

Zeins are grouped into α, β, γ, and δ types based on differences in aqueous solubility and the ability to form disulfide interactions, and the sequence of their encoding genes (Coleman and Larkins, 1999; Holding and Messing, 2013). Alpha zeins are the major component that accounts for more than 70% of all zein protein. Alpha zein proteins are encoded by four gene sub-families: *Z1A*, *Z1B*, *Z1C* and *Z1D*. *Z1A*, *Z1B* and *Z1D* sub-families encode the 19-kD α-zeins and 22-kD α-zeins are encoded by *Z1C* sub-familiy. The B73 reference line has 41 genes that encode alpha zeins and they are located on six chromosomal regions (Song et al., 2001; Song and Messing, 2002; Feng et al., 2009). A study comparing different maize backgrounds showed different gene copy numbers and expression levels (Jia et al., 2007; Xu and Messing, 2008; Feng et al., 2009; Miclaus et al.,

2011). For example, it was found that *Z1C* gene family contains 23 and 16 gene copies respectively in BSSS53 and B73. There were seven of 23 and six of 16 gene members from the *Z1C* family expressed, respectively, in BSSS53 and B73 at 18 DAP. These genes showed dramatic quantitative differences in expression patterns between the two backgrounds as measured by quantitative RT-PCR (Song et al., 2001; Song and Messing, 2002, 2003; Xu and Messing, 2008; Feng et al., 2009). Different from 19- and 22-kD α-zeins, the 50-, 27- and 16-kD γ-zein and the 15-kD β-zein are encoded by single genes. Relative to other genes that encode different zein fractions, γ-zein genes have earlier expression in endosperm, thus protein bodies start as small accretions consisting entirely of γ-zein (Woo et al., 2001). As protein bodies expand, α- and δ-zeins are sequestered into the core with the γ-zein remaining at the protein body periphery. 19-kD α-zeins are the most abundant class and the 22-kD α-zeins localize in between the central 19-kD α-zeins and the γ-zein peripheral shell (Figure. 2) (Holding et al., 2007; Guo et al., 2013).

Some maize mutants have opaque endosperms and they do not transmit light, whereas the endosperms of normal or wild type kernels are vitreous and translucent. Recessive mutations that cause an opaque kernel phenotype are generally termed "opaque" (*o1, o2, o5, o7, o9-o11, o13-o17*). Semi-dominant and dominant mutations are traditionally termed "floury", as exemplified by *floury1* (*fl1*) and *floury2* (*fl2*) (Gillikin et al., 1997; Gibbon and Larkins, 2005). Opaque and floury mutant kernels typically exhibit a starchy endosperm texture, low buoyant density and reduced kernel hardness. They often show defects in zein accumulation or in protein body formation (Lending and Larkins, 1992; Gillikin et al., 1997). Investigations of these mutants offer the opportunity to explore the

mechanisms that affect vitreous endosperm formation, which is essential for agronomic kernel traits in maize. Despite the importance of zeins, studies of certain opaque mutants show that factors in addition to zein synthesis or protein body packaging may also disrupt the vitreous endosperm and function as important determinants of kernel texture (Holding et al., 2010; Miclaus et al., 2011; Myers et al., 2011). These opaque mutants may have aberrations in processes such as amino acid biosynthesis, cytoskeletal function, or plastid development. Functional genomics methods, such as the one described in this thesis, will be useful tools to completely understand the mechanisms of maize endosperm development.

4 *Opaque 2* mutant and Quality Protein Maize (QPM)

Because protein deposits in cereal endosperm are usually deficient in the essential amino acids, lysine and tryptophan (Mertz et al., 1964), non-ruminant animals require amino acid supplementation if corn is the major component of their feed. It is often unfeasible to remedy the deficiency by protein supplementation, since it is costly and not realistic in the developing countries that rely on cereals as a main food resource. A primary objective in some plant breeding programs is to improve the cereal endosperm storage proteins by increasing the concentrations of the essential amino acids. A significant breakthrough to start the breeding for improved protein quality in maize was the isolation of *opaque-2* mutant. *Opaque2* (*o2*) is the most well studied opaque mutant, and the reduced accumulation of α zeins in *o2* results in increased non-zeins that collectively increase the lysine and tryptophan concentration (Mertz et al., 1964). The *O2* gene encodes a transcription factor that controls gene expression in endosperm, including 22-kD α-zeins

genes, and other genes such as pyruvate Pi dikinase (PPDK) that are not endosperm specific and thus result in pleiotropic effects in the mutant (Schmidt et al., 1990; Schmidt et al., 1992; Damerval and De Vienne, 1993; Maddaloni et al., 1996; Sofi et al., 2009). In *o2* kernels the 22-kD α-zeins accumulate much lower levels, which is thought to be a major cause of the opaque phenotype. In addition to natural mutations, studies show that silencing the 22-kD α-zeins genes by RNAi causes an opaque endosperm, which suggests that the reduced synthesis of the 22-kD α-zeins is sufficient to disrupt kernel hardness and vitreousness. The o2 mutation also affects proteins not directly regulated by the O2 transcription factor. For example, 19-kD α-zeins are decreased, though not eliminated in the mutant.

However, the low seed density and soft endosperm kernel phenotype of the *o2* mutant leads to increased mechanical damage when harvest or processing, greater susceptibility to diseases, and lower yields, which makes *o2* lines unsuitable for large scale commercial production, harvest, storage, or processing (Gibbon and Larkins, 2005). Subsequently, breeders restored the endosperm texture of *o2* to a normal hard and vitreous phenotype, while keeping its high quality of protein (Vasal et al., 1980). The resulting modified *o2*, which has gained broad popularity in developed countries, is known as Quality Protein Maize (QPM) (Prasanna et al., 2001). To date, QPM varieties and hybrids have been released in about 40 countries (Sofi et al., 2009). During QPM breeding, breeders monitored genotypes for *o2* mutant loci and lysine concentration in each generation, and selected *o2* mutants with vitreous endosperm. QPM retains the high levels of lysine and tryptophan because of the maintained decreased abundance in α-zeins. However, the

genetic basis of *o2* endosperm modification is poorly understood (Prasanna et al., 2001; Pixley and Bjarnason, 2002). The two to three-fold higher accumulation of the 27-kD γ-zein in QPM than in wild type and *o2* suggested that the degree of QPM endosperm vitreousness correlates closely with the amount of 27-kD γ-zein (Ortega and Bates, 1983; Wallace et al., 1990; Gibbon and Larkins, 2005). There are several research results supporting the correlation of 27-kD γ-zein level with endosperm modification (Geetha et al., 1991; Lopes and Larkins, 1991; Wu and Messing, 2010). First, a study showed that the 27-kD γ-zein gene co-locates on chromosome 7 with the most significant QTL for endosperm modification (Holding et al., 2008; Holding et al., 2011). Second, 27-kD γ-zein is believed to play an of importance role in protein body initiation (Lending and Larkins, 1989). Increased 27-kD γ-zein results in the accumulation of larger numbers of smaller protein bodies, which is believed to allow the formation of a glassy matrix endosperm in the same way that a lower number of larger protein bodies does in wild type endosperm (Dannenhoffer et al., 1995). Third, γ-zein was shown, using an RNAi approach, to be essential for endosperm modification in QPM (Wu et al., 2010).

5 Other maize opaque mutants

In addition to *opaque2*, several other opaque mutants have been characterized. Some of the mutations that cause the opaque phenotype occur directly in zein genes that have direct roles in prolamin protein body formation. These mutants include *Floury-2 (fl2), Defective endosperm B30 (De-B30),* and *Mucronate (Mc)* (Lending and Larkins, 1992; Gillikin et al., 1997). Some of the mutations are in genes that encode proteins that have indirect roles in prolamin protein body such as *Floury-1 (fl1)* and *Opaque-1 (o1)*

(Holding et al., 2007; Wang et al., 2012). Others are unrelated to prolamins and protein bodies, such as *zmAroDH-1 (mto140)*, *Opaque-5 (o5)* and *Opaque-7 (o7)* (Holding et al., 2010; Miclaus et al., 2011; Myers et al., 2011).

A number of unfolded protein response (UPR)-related genes like the ER binding protein (BiP) are elevated in all studied opaque mutants as a common feature (Fontes et al., 1991; Marocco et al., 1991; Hunter et al., 2002; Kim et al., 2006). BiP, an ER resident protein, is one member of the 70-kD heat shock protein family and has the property of recognizing and binding polypeptides that are incompletely assembled or unfolded, and continually produced during the normal protein synthesis (Gething, 1999). Studies in animal tissues show that abnormally folded proteins increase the level of BiP. The degree of abnormality in protein body morphology correlates with the increased level of BiP in mutant *fl2*, *Mc* and *De-B30* (Fontes et al., 1991; Marocco et al., 1991; Kim et al., 2006). These results suggest that the irregularity of protein deposition in ER causes an ER stress, and this endosperm ER stress, or a resulting energy crisis through increased ATP usage, may at least partially contribute to the disruption in formation of the vitreous endosperm (Bertolotti et al., 2000; Hunter et al., 2002; Guo et al., 2012).

The phenotypes of *fl2, De-B30* and *Mc* mutants result from the accumulation of dominant negative defective zeins. The resultant UPR in these mutants causes translational repression resulting in an overall decreased abundance of all zeins. Similarly to the *o2* mutant the relative proportion of non-zeins increases, which causes a similar, though less pronounced, increase in lysine concentration. The *fl2* mutant has a general decrease in all

zein accumulation, altered protein body morphology and a novel 24-kD α-zein storage protein that is visible on standard SDS-PAGE gels (Gillikin et al., 1997). In wild type, the 24-kD α-zein propeptide contains a signal peptide that would normally be removed by signal peptidase during protein synthesis and processing to produce the mature 22-kD α-zein. However, the *fl2* mutant results from a point mutation causing an Ala- to- Val substitution that blocks the signal peptidase cleavage and results in an unprocessed 24-kD α-zein. The 24-kD α-zeins with the abnormal signal peptide remain inappropriately tethered to the ER membrane (Gillikin et al., 1997).

De-B30 was first described in 1979 and the gene for *De-B30* was cloned in 2004 (Kim et al., 2004). Similarly to *fl2*, *De-B30* also results from a point mutation, causing a Pro substitution for Ser at the 15th amino acid in the signal peptide of 19-kD α-zeins. The amino acid substitution does not prevent signal peptide cleavage, but does affect the efficiency of the cleavage process (Kim et al., 2004). The uncleaved 19-kD α-zeins in *DeB30* are only detectable with western blotting, unlike the uncleaved 22-kD α-zeins of *fl2* that is readily detectable on a commassie blue SDS-PAGE gel. Thus the dominant negative effect in *DeB30* is the result of a very low level of abnormal protein. Both the 19- and 22-kD α-zeins are very hydrophobic proteins and are structurally very similar. These two points may explain why these two mutants share many phenotypic features. But why is *De-B30* a dominant mutant and *fl2* is a semi-dominant mutant? This could be because the *fl2* signal peptide may not be recognized by the signal peptidase at all, which results in the accumulation of precursor 24-kD α-zein. The signal peptidase complex may not remain on the protein precursor. Even though the *De-B30* signal peptide could not be

efficiently cleaved, it is recognized by signal peptidase. This may cause the signal peptidase complex to remain attached to the zein precursor, and not detaching as in *fl2*, which could cause a more severe and broader impact on metabolism in endosperm cells (Kim et al., 2004).

Similar to the maize mutants *fl2* and *De-B30*, a sorghum floury (non-vitreous/opaque) mutant (*P721Q*, Q for opaque) results from a missense mutation in the signal peptide of a sorghum prolamin kafirin, α-kafirin. We characterized this mutant in collaboration with Joachim Messing (Wu et al., 2013). Due to my significant involvement in this work, I present more detailed discussion of this mutant characterization in appendix 2.

The maize *Mc* mutant (Soave and Salamini, 1984) was identified as a novel, dominant opaque mutant. By comparing the mRNA profiles of Mc endosperm with that of nearly isogenic wild-type, a novel transcript in *Mc* were shown to correspond to a 16-kD γ-zein (Kim et al., 2006). Following cloning and sequencing of the mutant cDNA, a 38-nucleotide deletion was found that caused a frame shift mutation. This resulted in a nonsense sequence for the C-terminal 63 amino acids, which was terminated by a premature stop codon. Even though the transcript abundance of the 16-kD γ-zein gene is not affected by the frame shift mutation, the abundance of the 16-kD γ-zein protein is reduced in *Mc* compared to wild type, and the *Mc* mutant shows a general decreased abundance in all zein protein synthesis. In addition, protein bodies in *Mc* are misshapen, although not as much as in *fl2* and *De-B30* (Soave and Salamini, 1984; Kim et al., 2006).

Unlike *fl2, De-B30* and *Mc*, the phenotypes in opaque mutants such as *Floury-1 (fl1)* and *Opaque-1 (o1)* arise through mutation in non-zein encoding genes (Holding et al., 2007; Wang et al., 2012). *fl1* was first reported in 1912 and characterized and molecularly identified in 2007. *FL1* encodes a protein body ER-membrane specific protein (Holding et al., 2007). The gene encodes three predicted transmembrane domains and a domain of unknown function (DUF 593). *fl1* mutant has no change in the zein abundance, protein body shape or size. But the *fl1* null mutant does influence the 22-kD α-zein localization (Holding et al., 2007). On the contrary, the *o1* mutant has small and misshapen protein bodies and the ER is also abnormal and dilated. *O1* encodes the myosin XI protein that shares many structural features with known myosins (Wang et al., 2012). It is highly expressed in kernels and is associated with cisternal and protein body ER (Wang et al., 2012). A recent study in Arabidopsis showed that myosin can bind two previously uncharacterized transmembrane proteins called MyoB1/2 (Peremyslov et al., 2013). These two proteins share the common and conserved domain, DUF593, which was shown to act as the myosin receptor (Peremyslov et al., 2013). Since Fl1 is a protein body ER membrane specific protein and also contains DUF593 (Holding et al., 2007), this strongly suggests that F1 may function to attach protein bodies to myosin within the cytoskeleton. Though not yet demonstrated, O1 may have a direct or indirect functional interaction with Fl1 protein through DUF593.

The characterization of the above mutants shows that the size, structure and number of protein bodies affect the vitreous endosperm formation in the maize endosperm. However, mutations in mutant *mto140*, *o5* and *o7* that are not related to prolamins also result in

opaque phenotypes (Holding et al., 2010; Miclaus et al., 2011; Myers et al., 2011), which suggests that factors that are essential for vitreous endosperm formation are not necessarily directly related to prolamins. The characterization of *mto140* shows that its opaque phenotype co-segregates with a Mutator transposon insertion in an arogenate dehydrogenase gene *zmAroDH-1,* which is part of a four-member family in maize (Holding et al., 2010). Arogenate dehydrogenase plays an essential role in the synthesis of the amino acids tyrosine and phenylalanine from arogenate (Gaines et al., 1982; Connelly and Conn, 1985). The opaque phenotype in *mto140* results from a disruption in amino acid biosynthesis that causes a general disruption in protein synthesis (Holding et al., 2010). *o5* is specifically deficient in membrane galactolipids that are necessary for amyloplast and chloroplast function. This causes developmental defects and changes in starch production since the normal simple starch granules are replaced with compound granules separated by amyloplast membranes. The *O5* gene encodes the major biosynthetic enzyme MGD1 for synthesis of chloroplast membrane lipids (Myers et al., 2011). *o7* is another previously mapped maize starchy endosperm mutant, which were recently characterized at the molecular level. The *O7* gene encodes an acyl-activating enzyme and decrease amino acid biosynthesis possibly by affecting α-ketoglutaric acid and oxaloacetic acid (Miclaus et al., 2011; Wang et al., 2011). All mutants described above are listed in table 1.

6 Use of RNA interference for improving grain protein quality

In addition to naturally occurring opaque mutations, quantitatively modifying zein abundance by RNA interference (RNAi) is another method to improve the protein quality

of the maize seed and to study zein function as described in appendix 1. Increase in lysine can be achieved by modifying the activity or regulatory properties of the enzymes involved in lysine catabolism (Capell and Christou, 2004). RNAi has also been used as an effective tool to substantially reduce or eliminate the synthesis of entire subfamilies of zeins, and then to study the functions of specific zein classes. The recessive zein mutants are lack since multiple members of the α-zein subfamilies have a high-level expression, which results in substantial functional redundancy, masking the effects of recessive mutants in individual family members. Also, naturally occurring opaque mutants often carry not only positive nutritional benefits but also negative pleiotropic traits. RNAi is an effective genetic tool to create dominant lines with improved protein quality that are not afflicted by some of the negative pleotropic affects seen in mutants such as *o2*. One example is the decreased of PPDK, a central glycolytic enzyme that is regulated by *O2*.

α-zeins, as the main zein components in maize endosperm, are the major determinants of protein quality. Like most cereal prolamins, zeins are deficient in lysine and tryptophan, and the dominant abundance of prolamins imparts poor overall protein quality. Kernels with reduced zein and increased non-zein have higher levels of lysine and tryptophan (Habben et al., 1993). The two subfamilies of α-zeins, 22-kD α-zeins and 19-kD α-zeins, comprise more than 20% and 40% of the total zein, respectively (Huang et al., 2006). As mentioned above, the naturally occurring *o2* mutant has improved nutritional quality since the *O2* gene regulates the expression of a subset of 22-kD α-zeins. But the *o2* mutation is recessive and easily lost by wild type pollen contamination during crossing prior to selection for vitreous lines in QPM breeding. O2 plays an important role in

multiple endosperm metabolic pathways and regulates genes other than 22-kD α-zeins. The *o2* mutation therefore causes pleiotropic effects in a variety of endosperm metabolic pathways. However, several 22-kD α-zeins are not under the control of O2, which means that not all 22-kD α-zeins are directly repressed in the *o2* mutant. Although in many cases, reduction of gene expression by RNAi is not complete and may not produce a visible phenotype, RNAi was used to create dominant mutations that suppress zein genes without interrupting O2 synthesis (Segal et al., 2003; Huang et al., 2004; Huang et al., 2006; Wu and Messing, 2010; Guo et al., 2013). Constructs derived from 22-kD α-zein genes were shown to reduce the α-zein fraction considerably and also produce an opaque phenotype and higher level of lysine (Segal et al., 2003). This was the first study to demonstrate the separation of the effects on the transcription of zein genes from other genes and specifically repressed the synthesis of 22-kD α-zeins in the presence of normal *O2* expression. The transgenic line had protein bodies of reduced size with a distorted lobed appearance. The study also showed that a sufficient amount of the 22-kD α-zein is necessary for a normal protein body shape. A later study demonstrated that the targeted reduction of the other α-zeins component, the 19-kD α-zeins, also results in the increase of lysine and tryptophan concentration (Huang et al., 2004). Transgenic maize kernels engineered to reduce either 22- or 19-kD α-zeins have a moderate increase in lysine level. Reducing both the 22- and 19-kD subfamilies of α-zeins further enhanced lysine and tryptophan concentration by a greater reduction in zein protein accumulation and an increase in lysine-rich, non-zein proteins (Huang et al., 2006).

In sorghum, transgenic plants were generated with targeted reduction of kernel storage protein kafirin accumulation (Kumar et al., 2012). Down-regulation of a predicted 29-kD α-kafirin led to higher accumulation of non-kafirin proteins, which resulted in increased lysine concentration. Indigestibility of sorghum grain is a major limitation for its use as a human and animal food. The transgenic line with down-regulated 29-kD α-kafirin had enhanced protein digestibility in both raw and cooked samples. The RNAi line for the predicted 29-kD α-kafirin alters the morphology of protein bodies, which is similar to the above-mentioned 22-kD α-zein RNAi lines and dominant 22-kD α-zein signal peptide mutant *fl2*. In addition, RNAi technology was also been used to reduce accumulation of γ-zeins, and to obtain insight into their redundant and non-redundant functions (Guo et al., 2013). 27- and 16-kD γ-zeins share sequence similarity. 27-/16-kD γ-zeins RNAi and 15-kD β-zein RNAi events separately caused minor morphological changes to protein bodies. Combined γ-zeins RNAi and β-zein RNAi resulted in significant protein body morphological changes (Guo et al., 2013).

7 Our contributions to knowledge of maize opaque mutants and QPM

Given the potential benefits of QPM, continued research on the genetic and biochemical mechanisms that are responsible for altered kernel texture and protein quality is of high importance. The identification and investigation of other opaque mutants will help provide clues to better understand how kernel texture is influenced in maize (Gibbon and Larkins, 2005). To generate new mutants that will increase our understanding of the factors that contribute to the kernel vitreousness, we conducted mutagenesis and screening for starchy, opaque kernel mutants.

7.1 Methods for creating mutants: EMS mutagenesis, transposon insertions and irradiation

A variety of resources for forward and reverse genetics exist in maize and these are based on EMS mutagenesis and transposon insertions (Settles, 2005; Weil et al., 2005). EMS mutagenesis creates mostly point mutations that often result in knock-down alleles, which are useful in studying gene function. Transposons that create a high frequency of knock-out alleles have been used for a variety of mutagenesis schemes (McCarty et al., 2005; Settles et al., 2007; Vollbrecht et al., 2010) and have led to various reverse genetics resources. Though transposon insertions can create null alleles, these can be leaky mutations, leading to weak phenotypes. Physical deletion of genes through irradiation is one way to remove this ambiguity. Radiation has been used to induce physical deletions of genes in Arabidopsis and this has been shown to be useful for generating mutant populations. Such populations can be set up for reverse genetics using PCR based screens (Li et al., 2001). The chromosomal alterations in Arabidopsis included point mutations, translocations, insertions, small deletions and large deletions ranging in size from kilobases to several megabases. Deletion mutagenesis in maize has potential as a tool for investigating a number of biological processes such as endosperm maturation and other aspects of seed development. Even without saturation mutagenesis, an indexed collection of single and multiple gene deletions would be useful in conjunction with existing forward and reverse genetics resources.

The potential of γ-radiation for reverse genetics has been investigated (Wu et al., 2005;

Sato et al., 2006) and although the mutagenesis rate is lower than with EMS mutagenesis, γ-radiation results in a higher proportion of knockout mutations (Sato et al., 2006). The dosage of radiation must be optimized to maximize the mutagenesis rate while allowing an acceptable level of viability of the M1 plants. Low and moderate radiation doses cause a relatively high proportion of useful mutants with normal yield (Dumanovc et al., 1969), while high radiation doses induce a greater proportion of large deletions that remove genes that are essential for viability (Naito et al., 2005). In an Arabidopsis irradiation experiment, the majority of large deletions was not transmitted from the M1 to the M2 generation and was assumed to result from loss of genes required for seed germination, gamete development and other aspects of reproductive development (Naito et al., 2005). However, in large genomes with lower gene density, the tolerance for large deletion could be higher. In pilot studies with QPM maize seed, we observed a 50% germination rate using a 20krad dose. This gives the highest mutation rate while still having enough viable lines to screen for mutants of the specific phenotype. This dosage was subsequently used in our mutagenesis project.

7.2 K0326Y QPM mutagenesis population

The K0326Y inbred line is one of the Quality Protein Maize (QPM) developed in South Africa. To identify *o2* modifier genes and to investigate deletion mutagenesis, we used γ-irradiation combined with exon-capture DNA sequencing and RNA-seq as a maize functional genomics tool. We created a K0326Y QPM mutagenesis population and identified the opaque variants (Yuan et al., 2014). In the K0326Y QPM mutagenesis project, around 2000 K0326Y seeds were irradiated, which were termed the M1

generation. 293 ears were recovered in M2 generation. Twenty kernels for all M2 ears were propagated in field. This enabled heritability testing of dominant M2 phenotypes, rescue of M2 ears with few kernels and identification of recessive mutants appearing in the M3 generation. In addition to *o2*-modifier genes, this approach also potentially allows functional genomic dissection of genes more generally involved in endosperm formation as well as seed and plant development.

In the M3 generation, eight lines exhibited segregation for vegetative plant phenotype, five of which were obviously shorter that non-mutagenized K02326Y. Notably, line 43 segregated for dwarf plants. Three lines segregated for partial chlorosis or were albino during seedling stage and did not survive to reproduction. 27 lines exhibited clear segregation for opaque kernel phenotypes. Of these 27 lines, opaque kernels in 17 lines were not viable and were presumed to be pleiotropic mutations. The other ten lines had fully viable opaque kernels and thus were candidates for factors specifically involved in *o2* endosperm modification and/or vitreous endosperm formation in general. We conducted exon-seq for line 43, and the other opaque kernel mutants including line 44, 107, 112, 121, 133 and 198. Part of the data (lines 107 and 198) has been analyzed, and in the case of line 107, has enabled us to identify the specific mutation causing the phenotype (Yuan et al., 2014). We found genome-wide large deletions by γ ray mutagenesis in non-reference maize line can be identified through a combination of assembly of Illumina reads of the K0326Y exome against the B73 genome and integration of RNA-seq data. We have already conducted Bulk Segregant Analysis (BSA) to map the mutations in lines 44, 112, 121 and 198. Additional BSA mapping populations

for the rest of the opaque mutants will be developed in fall 2014. We are characterizing more mutants by combining exon-seq, RNA-seq and BSA mapping. Of all these mutants, line 107 and line 198 are the two best-characterized mutants. Detailed characterizations of line 107 and line 198 are shown in Chapters 2 and 3 respectively.

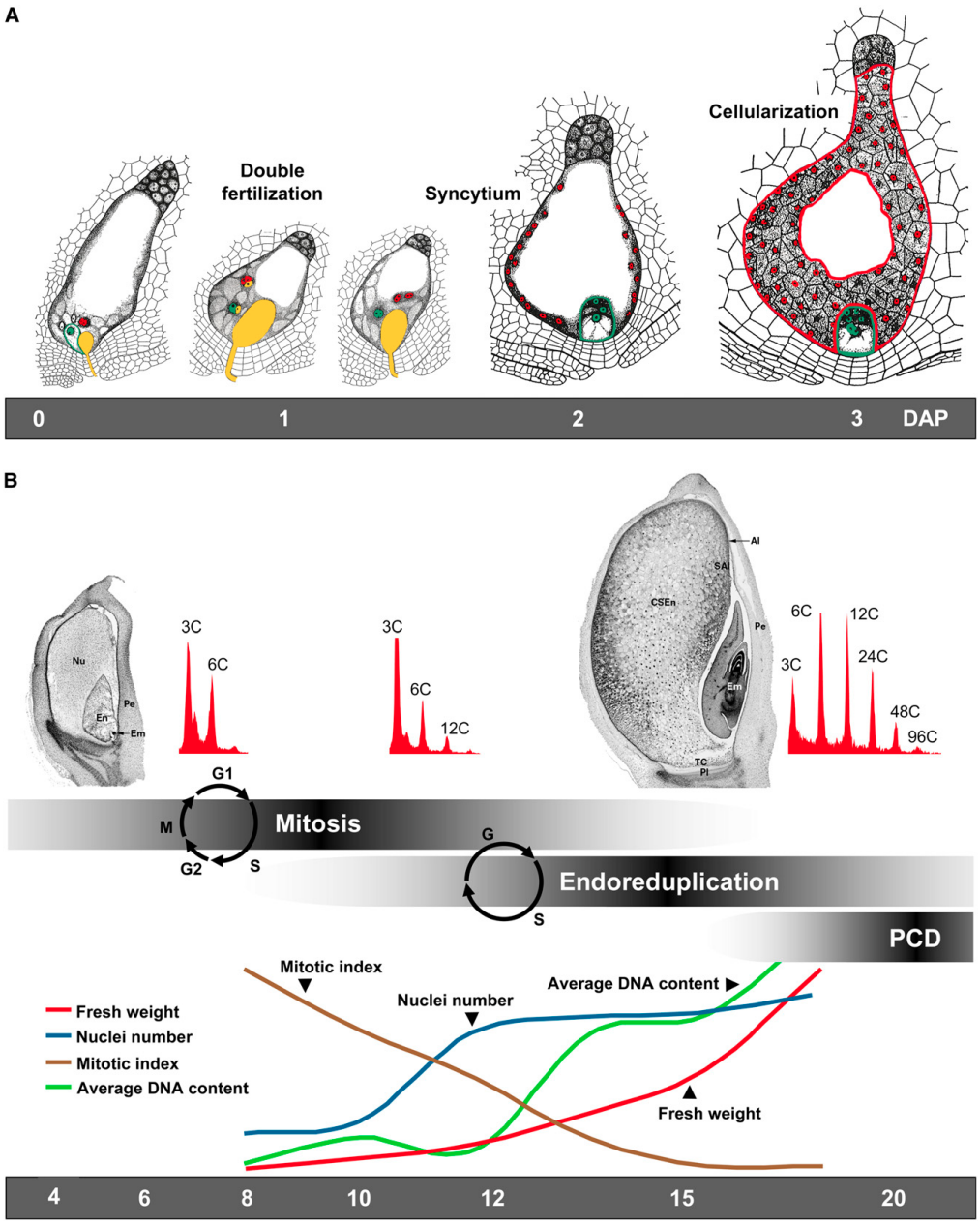


Figure 1. Phases in maize endosperm development. (Source: Sabelli and Larkins, 2009)
A, Double fertilization, syncytium formation, and cellularization. Yellow: pollen tube and sperm nuclei; Red: polar nuclei; Green: egg cell nucleus and embryo nuclei. Outlines of the multicellular endosperm and embryo are drawn in red and green, respectively.
B, Mitotic cell proliferation, endoreduplication and PCD. Red line: fresh weight; Blue line: nuclei number; Brown line: mitotic index; Green line: average DNA content.
Al, Aleurone; CSEn, central starchy endosperm; Em, embryo; En, endosperm; Nu, nucellus; Pe, pericarp; Pl, placentochalaza; SAl, subaleurone layer; TC, transfer cells.

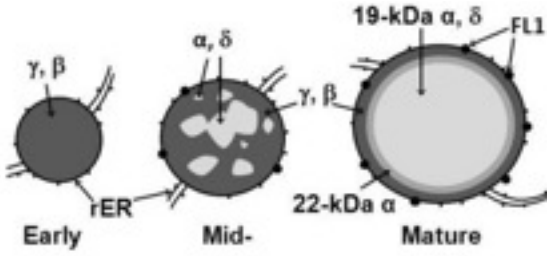


Figure 2. Current model for zein protein body architecture at early, mid- and mature stages. (Source: Guo et al., 2013) Protein bodies first appear as small γ-zein accretions (left), and later, α- and δ-zein penetrate and appear as inclusions with the γ-zein periphery (middle). Mature protein bodies have an even, round shape, with α- and δ-zeins confined within the core surrounded by a shell of γ-and β-zeins (right). FL1 is a non-zein protein body-ER membrane specific protein that participates in the zein organization (Holding et al,. 2007).

Table 1. Opaque mutants described

Mutants	Defect
o2	Loss of a transcription factor that regulates including 22-kD α-zeins
fl2	Uncleaved 22-kD α-zeins
De-B30	Uncleaved 19-kD α-zeins
Mc	Nonsense 16-kD γ-zeins
fl1	Deficient protein body ER-membrane specific protein
o1	Deficient myosine XI
mto140	Deficient arogenate dehydrogenase and disruption in amino acid biosynthesis
o5	Deficient membrane galactolipids
o7	Deficient acyl-activating enzyme and disruption in amino acid biosynthesis

REFERENCES

Becraft PW, Yi G (2010) Regulation of aleurone development in cereal grains. Journal of experimental botany: erq372

Bertolotti A, Zhang Y, Hendershot LM, Harding HP, Ron D (2000) Dynamic interaction of BiP and ER stress transducers in the unfolded-protein response. Nature cell biology 2: 326-332

Bolon Y-T, Haun WJ, Xu WW, Grant D, Stacey MG, Nelson RT, Gerhardt DJ, Jeddeloh JA, Stacey G, Muehlbauer GJ (2011) Phenotypic and genomic analyses of a fast neutron mutant population resource in soybean. Plant physiology 156: 240-253

Bressani R, Rios BJ (1962) The chemical and essential amino acid composition of twenty-five selections of grain sorghum. Cereal Chem 39: 50-58

Buttrose M (1963) Ultrastructure of the developing aleurone cells of wheat grain. Australian Journal of Biological Sciences 16: 768-774

Capell T, Christou P (2004) Progress in plant metabolic engineering. Current Opinion in Biotechnology 15: 148-154

Charlton W, Keen C, Merriman C, Lynch P, Greenland A, Dickinson H (1995) Endosperm development in Zea mays; implication of gametic imprinting and paternal excess in regulation of transfer layer development. Development 121: 3089-3097

Coleman CE, Larkins BA (1999) The prolamins of maize. *In* seed Proteins. Springer, pp 109-139

Connelly J, Conn E (1985) Tyrosine biosynthesis in Sorghum bicolor: isolation and regulatory properties of arogenate dehydrogenase. Zeitschrift fur Naturforschung. C, Journal of biosciences 41: 69-78

Damerval C, De Vienne D (1993) Quantification of dominance for proteins pleiotropically affected by opaque-2 in maize. HEREDITY-LONDON- 70: 38-38

Dannenhoffer JM, Bostwick DE, Or E, Larkins BA (1995) opaque-15, a maize mutation with properties of a defective *opaque-2* modifier. Proceedings of the National Academy of Sciences 92: 1931-1935

Dumanovc J, Denic M, Ehrenberg L, Bergstrand K (1969) Radiation-induced heritable variation of quantitative characters in wheat. Hereditas 62: 221-238

Duodu K, Taylor J, Belton P, Hamaker B (2003) Factors affecting sorghum protein digestibility. Journal of Cereal Science 38: 117-131

Esen A (1987) A proposed nomenclature for the alcohol-soluble proteins (zeins) of maize. Journal of Cereal Science 5: 117-128

Feng L, Zhu J, Wang G, Tang Y, Chen H, Jin W, Wang F, Mei B, Xu Z, Song R (2009) Expressional profiling study revealed unique expressional patterns and dramatic expressional divergence of maize α-zein super gene family. Plant molecular biology 69: 649-659

Fontes E, Shank BB, Wrobel RL, Moose SP, OBrian GR, Wurtzel ET, Boston RS (1991) Characterization of an immunoglobulin binding protein homolog in the maize floury-2 endosperm mutant. The Plant Cell Online 3: 483-496

Gaines C, Byng G, Whitaker R, Jensen R (1982) L-Tyrosine regulation and biosynthesis via arogenate dehydrogenase in suspension-cultured cells of Nicotiana silvestris Speg. et Comes. Planta 156: 233-240

Geetha K, Lending CR, Lopes MA, Wallace JC, Larkins BA (1991) opaque-2 modifiers increase gamma-zein synthesis and alter its spatial distribution in maize endosperm. The Plant Cell Online 3: 1207-1219

Gething M-J (1999) Role and regulation of the ER chaperone BiP. *In* Seminars in cell & developmental biology, Vol 10. Elsevier, pp 465-472

Gibbon BC, Larkins BA (2005) Molecular genetic approaches to developing quality protein maize. Trends Genet 21: 227-233

Gillikin JW, Zhang F, Coleman CE, Bass HW, Larkins BA, Boston RS (1997) A defective signal peptide tethers the floury-2 zein to the endoplasmic reticulum membrane. Plant physiology 114: 345-352

Guo X, Ronhovde K, Yuan L, Yao B, Soundararajan MP, Elthon T, Zhang C, Holding DR (2012) Pyrophosphate-dependent fructose-6-phosphate 1-phosphotransferase induction and attenuation of Hsp gene expression during endosperm modification in quality protein maize. Plant physiology 158: 917-929

Guo X, Yuan L, Chen H, Sato SJ, Clemente TE, Holding DR (2013) Nonredundant function of zeins and their correct stoichiometric ratio drive protein body formation in maize endosperm. Plant Physiol 162: 1359-1369

Habben JE, Kirleis AW, Larkins BA (1993) The origin of lysine-containing proteins in opaque-2 maize endosperm. Plant molecular biology 23: 825-838

Hamaker B, Kirleis A, Butler L, Axtell J, Mertz E (1987) Improving the in vitro protein digestibility of sorghum with reducing agents. Proceedings of the National Academy of Sciences 84: 626-628

Hamaker B, Mertz E, Axtell J (1994) Effect of extrusion on sorghum kafirin solubility. Cereal chemistry 71: 515-515

Hannah LC (2007) Starch formation in the cereal endosperm. *In* Endosperm. Springer, pp 179-193

Holding DR (2014) Recent advances in the study of prolamin storage protein organization and function. Plant Evolution and Development

Holding D, Larkins B (2006) The development and importance of zein protein bodies in maize endosperm. Maydica 51: 243

Holding D, Messing J (2013) Evolution, structure, and function of prolamin storage proteins. Seed Genomics. John Wiley & Sons, New York: 139-158

Holding DR, Hunter BG, Chung T, Gibbon BC, Ford CF, Bharti AK, Messing J, Hamaker BR, Larkins BA (2008) Genetic analysis of opaque2 modifier loci in quality protein maize. Theoretical and Applied Genetics 117: 157-170

Holding DR, Hunter BG, Klingler JP, Wu S, Guo X, Gibbon BC, Wu R, Schulze J-M, Jung R, Larkins BA (2011) Characterization of opaque2 modifier QTLs and candidate genes in recombinant inbred lines derived from the K0326Y quality protein maize inbred. Theoretical and applied genetics 122: 783-794

Holding DR, Meeley RB, Hazebroek J, Selinger D, Gruis F, Jung R, Larkins BA (2010) Identification and characterization of the maize arogenate dehydrogenase gene family. J Exp Bot 61: 3663-3673

Holding DR, Otegui MS, Li B, Meeley RB, Dam T, Hunter BG, Jung R, Larkins BA (2007) The maize floury1 gene encodes a novel endoplasmic reticulum protein involved in zein protein body formation. The Plant Cell Online 19: 2569-2582

Huang S, Adams WR, Zhou Q, Malloy KP, Voyles DA, Anthony J, Kriz AL, Luethy MH (2004) Improving nutritional quality of maize proteins by expressing sense and antisense zein genes. Journal of agricultural and food chemistry 52: 1958-1964

Huang S, Frizzi A, Florida CA, Kruger DE, Luethy MH (2006) High lysine and high tryptophan transgenic maize resulting from the reduction of both 19- and 22-kD alpha-zeins. Plant Mol Biol 61: 525-535

Hunt C, Kezar W, Vinande R (1989) Yield, chemical composition and ruminal fermentability of corn whole plant, ear, and stover as affected by maturity. Journal of production agriculture 2: 357-361

Hunter BG, Beatty MK, Singletary GW, Hamaker BR, Dilkes BP, Larkins BA, Jung R (2002) Maize opaque endosperm mutations create extensive changes in patterns of gene expression. The Plant Cell Online 14: 2591-2612

James MG, Denyer K, Myers AM (2003) Starch synthesis in the cereal endosperm. Current Opinion in Plant Biology 6: 215-222

Jia H, Nettleton D, Peterson JM, Vazquez-Carrillo G, Jannink J-L, Scott MP (2007) Comparison of transcript profiles in wild-type and maize endosperm in different genetic backgrounds. Crop science 47: S-45-S-59

Johri M, Coe Jr E (1983) Clonal analysis of corn plant development: I. The development of the tassel and the ear shoot. Developmental biology 97: 154-172

Kiesselbach T (1949) The Structure and Reproduction of Corn, research bulletin No. 161. University of Nebraska Agricultura1 Experiment Station, Lincoln, NE: 61-84

Kiesselbach TA (1999) The structure and reproduction of corn. Cold spring harbor laboratory press

Kim CS, Gibbon BC, Gillikin JW, Larkins BA, Boston RS, Jung R (2006) The maize Mucronate mutation is a deletion in the 16-kDa gamma-zein gene that induces the unfolded protein response. Plant J 48: 440-451

Kim CS, Hunter BG, Kraft J, Boston RS, Yans S, Jung R, Larkins BA (2004) A defective signal peptide in a 19-kD alpha-zein protein causes the unfolded protein response and an opaque endosperm phenotype in the maize De*-B30 mutant. Plant Physiol 134: 380-387

Kim CS, Woo Y-m, Clore AM, Burnett RJ, Carneiro NP, Larkins BA (2002) Zein protein interactions, rather than the asymmetric distribution of zein mRNAs on endoplasmic reticulum membranes, influence protein body formation in maize endosperm. The Plant Cell Online 14: 655-672

Kranz E, von Wiegen P, Quader H, Lörz H (1998) Endosperm development after fusion of isolated, single maize sperm and central cells in vitro. The Plant Cell Online 10: 511-524

Kumar T, Dweikat I, Sato S, Ge Z, Nersesian N, Chen H, Elthon T, Bean S, Ioerger BP, Tilley M, Clemente T (2012) Modulation of kernel storage proteins in grain sorghum (Sorghum bicolor (L.) Moench). Plant Biotechnol J 10: 533-544

Lending C, Larkins B (1992) Effect of the floury-2 locus on protein body formation during maize endosperm development. Protoplasma 171: 123-133

Lending CR, Larkins BA (1989) Changes in the zein composition of protein bodies during maize endosperm development. The Plant Cell Online 1: 1011-1023

Li G, Wang D, Yang R, Logan K, Chen H, Zhang S, Skaggs MI, Lloyd A, Burnett WJ, Laurie JD, Hunter BG, Dannenhoffer JM, Larkins BA, Drews GN, Wang X, Yadegari R (2014) Temporal patterns of gene expression in developing maize endosperm identified through transcriptome sequencing. Proc Natl Acad Sci U S A

Li X, Song Y, Century K, Straight S, Ronald P, Dong X, Lassner M, Zhang Y (2001) A fast neutron deletion mutagenesis-based reverse genetics system for plants. The Plant Journal 27: 235-242

Liu S, Chen HD, Makarevitch I, Shirmer R, Emrich SJ, Dietrich CR, Barbazuk WB, Springer NM, Schnable PS (2010) High-throughput genetic mapping of mutants via quantitative single nucleotide polymorphism typing. Genetics 184: 19-26

Lopes MA, Larkins BA (1991) Gamma-Zein Content is Ralated to Endosperm Modification in Quality Protein Maize. Crop science 31: 1655-1662

MacLeon Jr W, Lopez dRG, Placko RP, Graham GG (1981) Protein quality and digestibility of sorghum in preschool children: balance studies and plasma free amino acids. The Journal of nutrition 111: 1928-1936

Maddaloni M, Donini G, Balconi C, Rizzi E, Gallusci P, Forlani F, Lohmer S, Thompson R, Salamini F, Motto M (1996) The transcriptional activatorOpaque-2 controls the expression of a cytosolic form of pyruvate orthophosphate dikinase-1 in maize endosperms. Molecular and General Genetics MGG 250: 647-654

Marocco A, Santucci A, Cerioli S, Motto M, Di Fonzo N, Thompson R, Salamini F (1991) Three high-lysine mutations control the level of ATP-binding HSP70-like proteins in the maize endosperm. The Plant Cell Online 3: 507-515

McCarty DR, Mark Settles A, Suzuki M, Tan BC, Latshaw S, Porch T, Robin K, Baier J, Avigne W, Lai J (2005) Steady-state transposon mutagenesis in inbred maize. The Plant Journal 44: 52-61

Mertz ET, Bates LS, Nelson OE (1964) Mutant gene that changes protein composition and increases lysine content of maize endosperm. Science 145: 279-280

Mertz ET, Hassen MM, Cairns-Whittern C, Kirleis AW, Tu L, Axtell JD (1984) Pepsin digestibility of proteins in sorghum and other major cereals. Proceedings of the National Academy of Sciences 81: 1-2

Miclaus M, Wu Y, Xu J-H, Dooner HK, Messing J (2011) The maize high-lysine mutant opaque7 is defective in an acyl-CoA synthetase-like protein. Genetics 189: 1271-1280

Miclaus M, Xu J-H, Messing J (2011) Differential gene expression and epiregulation of alpha zein gene copies in maize haplotypes. PLoS genetics 7: e1002131

Myers AM, James MG, Lin Q, Yi G, Stinard PS, Hennen-Bierwagen TA, Becraft PW (2011) Maize opaque5 encodes monogalactosyldiacylglycerol synthase and specifically affects galactolipids necessary for amyloplast and chloroplast function. The Plant Cell Online 23: 2331-2347

Naito K, Kusaba M, Shikazono N, Takano T, Tanaka A, Tanisaka T, Nishimura M (2005) Transmissible and nontransmissible mutations induced by irradiating Arabidopsis thaliana pollen with γ-rays and carbon ions. Genetics 169: 881-889

Olsen O-A (2001) Endosperm development: cellularization and cell fate specification. Annual review of plant biology 52: 233-267

Olsen O-A, Linnestad C, Nichols SE (1999) Developmental biology of the cereal endosperm. Trends in plant science 4: 253-257

Oria MP, Hamaker BR, Axtell JD, Huang C-P (2000) A highly digestible sorghum mutant cultivar exhibits a unique folded structure of endosperm protein bodies. Proceedings of the National Academy of Sciences 97: 5065-5070

Ortega E, Bates L (1983) Biochemical and agronomic studies of two modified hard-endosperm opaque-2 maize (Zea mays L.) populations [Cultivars, protein components]. Cereal Chemistry 60

Peremyslov VV, Morgun EA, Kurth EG, Makarova KS, Koonin EV, Dolja VV (2013) Identification of myosin XI receptors in Arabidopsis defines a distinct class of transport vesicles. Plant Cell 25: 3022-3038

Peremyslov VV, Morgun EA, Kurth EG, Makarova KS, Koonin EV, Dolja VV (2013) Identification of myosin XI receptors in Arabidopsis defines a distinct class of transport vesicles. The Plant Cell Online 25: 3022-3038

Pixley KV, Bjarnason MS (2002) Stability of grain yield, endosperm modification, and protein quality of hybrid and open-pollinated quality protein maize (QPM) cultivars. Crop Science 42: 1882-1890

Poethig R, Coe Jr E, Johri M (1986) Cell lineage patterns in maize embryogenesis: a clonal analysis. Developmental Biology 117: 392-404

Prasanna B, Vasal S, Kassahun B, Singh N (2001) Quality protein maize. CURRENT SCIENCE-BANGALORE- 81: 1308-1319

Randolph LF (1936) Developmental morphology of the caryopsis in maize. United States Department of Agriculture

Sabelli PA, Larkins BA (2009) The development of endosperm in grasses. Plant Physiol 149: 14-26

Sabelli PA, Liu Y, Dante RA, Lizarraga LE, Nguyen HN, Brown SW, Klingler JP, Yu J, LaBrant E, Layton TM, Feldman M, Larkins BA (2013) Control of cell proliferation, endoreduplication, cell size, and cell death by the retinoblastoma-related pathway in maize endosperm. Proc Natl Acad Sci U S A 110: E1827-1836

Sato Y, Shirasawa K, Takahashi Y, Nishimura M, Nishio T (2006) Mutant selection from progeny of gamma-ray-irradiated rice by DNA heteroduplex cleavage using Brassica petiole extract. Breeding science 56: 179-183

Schmidt RJ, Burr FA, Aukerman MJ, Burr B (1990) Maize regulatory gene opaque-2 encodes a protein with a" leucine-zipper" motif that binds to zein DNA. Proceedings of the National Academy of Sciences 87: 46-50

Schmidt RJ, Ketudat M, Aukerman MJ, Hoschek G (1992) Opaque-2 is a transcriptional activator that recognizes a specific target site in 22-kD zein genes. The Plant Cell Online 4: 689-700

Segal G, Song R, Messing J (2003) A new opaque variant of maize by a single dominant RNA-interference-inducing transgene. Genetics 165: 387-397

Settles A (2005) Maize community resources for forward and reverse genetics. Maydica 50: 405

Settles AM, Holding DR, Tan BC, Latshaw SP, Liu J, Suzuki M, Li L, O'Brien BA, Fajardo DS, Wroclawska E (2007) Sequence-indexed mutations in maize using the UniformMu transposon-tagging population. BMC genomics 8: 116

Shull JM, Watterson JJ, Kirleis A (1992) Purification and immunocytochemical localization of kafirins inSorghum bicolor (L. Moench) endosperm. Protoplasma 171: 64-74

Shull JM, Watterson JJ, Kirleis AW (1991) Proposed nomenclature for the alcohol-soluble proteins (kafirins) of Sorghum bicolor (L. Moench) based on molecular weight, solubility, and structure. Journal of Agricultural and Food Chemistry 39: 83-87

Smith AM (1999) Making starch. Current opinion in plant biology 2: 223-229

Soave C, Salamini F (1984) Organization and regulation of zein genes in maize endosperm. Philosophical Transactions of the Royal Society of London. B, Biological Sciences 304: 341-347

Sofi P, Wani SA, Rather A, Wani SH (2009) Review article: Quality protein maize (QPM): Genetic.

Song R, Llaca V, Linton E, Messing J (2001) Sequence, regulation, and evolution of the maize 22-kD α zein gene family. Genome research 11: 1817-1825

Song R, Messing J (2002) Contiguous genomic DNA sequence comprising the 19-kD zein gene family from maize. Plant physiology 130: 1626-1635

Song R, Messing J (2003) Gene expression of a gene family in maize based on noncollinear haplotypes. Proceedings of the National Academy of Sciences 100: 9055-9060

Song R, Segal G, Messing J (2004) Expression of the sorghum 10-member kafirin gene cluster in maize endosperm. Nucleic acids research 32: e189-e189

Tsai C, Huber D, Warren H (1978) Relationship of the kernel sink for N to maize productivity. Crop science 18: 399-404

Vasal S, Villegas E, Bjarnason M, Gelaw B, Goertz P (1980) Genetic modifiers and breeding strategies in developing hard endosperm opaque-2 materials. Improvement of quality traits of maize for grain and silage use.: 37-73

Vollbrecht E, Duvick J, Schares JP, Ahern KR, Deewatthanawong P, Xu L, Conrad LJ, Kikuchi K, Kubinec TA, Hall BD (2010) Genome-wide distribution of transposed Dissociation elements in maize. The Plant Cell Online 22: 1667-1685

Wallace JC, Lopes MA, Paiva E, Larkins BA (1990) New methods for extraction and quantitation of zeins reveal a high content of γ-zein in modified opaque-2 maize. Plant physiology 92: 191-196

Wang G, Sun X, Wang G, Wang F, Gao Q, Sun X, Tang Y, Chang C, Lai J, Zhu L, Xu Z, Song R (2011) Opaque7 encodes an acyl-activating enzyme-like protein that affects storage protein synthesis in maize endosperm. Genetics 189: 1281-1295

Wang G, Wang F, Wang G, Wang F, Zhang X, Zhong M, Zhang J, Lin D, Tang Y, Xu Z (2012) Opaque1 encodes a myosin XI motor protein that is required for endoplasmic reticulum motility and protein body formation in maize endosperm. The Plant Cell Online 24: 3447-3462

Weil C, Monde R, Till R, Comai L, Henikoff S (2005) Mutagenesis and functional genomics in maize. Maydica 50: 415

Wong JH, Lau T, Cai N, Singh J, Pedersen JF, Vensel WH, Hurkman WJ, Wilson JD, Lemaux PG, Buchanan BB (2009) Digestibility of protein and starch from sorghum is linked to biochemical and structural features of grain endosperm. Journal of cereal science 49: 73-82

Woo Y-M, Hu DW-N, Larkins BA, Jung R (2001) Genomics analysis of genes expressed in maize endosperm identifies novel seed proteins and clarifies patterns of zein gene expression. The Plant Cell Online 13: 2297-2317

Wu J-L, Wu C, Lei C, Baraoidan M, Bordeos A, Madamba M, Suzette R, Ramos-Pamplona M, Mauleon R, Portugal A (2005) Chemical-and irradiation-induced mutants of indica rice IR64 for forward and reverse genetics. Plant molecular biology 59: 85-97

Wu Y, Holding DR, Messing J (2010) γ-Zeins are essential for endosperm modification in quality protein maize. Proceedings of the National Academy of Sciences 107: 12810-12815

Wu Y, Messing J (2010) RNA interference-mediated change in protein body morphology and seed opacity through loss of different zein proteins. Plant Physiol 153: 337-347

Wu Y, Yuan L, Guo X, Holding DR, Messing J (2013) Mutation in the seed storage protein kafirin creates a high-value food trait in sorghum. Nature communications 4

Xu J-H, Bennetzen JL, Messing J (2012) Dynamic gene copy number variation in collinear regions of grass genomes. Molecular biology and evolution 29: 861-871

Xu J-H, Messing J (2008) Organization of the prolamin gene family provides insight into the evolution of the maize genome and gene duplications in grass species. Proceedings of the National Academy of Sciences 105: 14330-14335

Yuan L, Dou Y, Kianian SF, Zhang C, Holding DR (2014) Deletion mutagenesis identifies a haploinsufficient role for gamma-zein in opaque2 endosperm modification. Plant Physiol 164: 119-130

CHAPTER 2

DELETION MUTAGENESIS IDENTIFIES A HAPLOINSUFFICIENT ROLE FOR GAMMA-ZEIN IN OPAQUE 2 ENDOSPERM MODIFICATION (Yuan et al., 2014)

ABSTRACT

Quality Protein Maize (QPM) is a hard kernel variant of the high-lysine mutant, *opaque-2*. Using γ-irradiation, we created opaque QPM variants to identify *opaque-2* modifier genes and to investigate deletion mutagenesis combined with Illumina sequencing as a maize functional genomics tool. A K0326Y-QPM deletion mutant (line 107) was null for the 27- and 50-kD γ-zeins and abolished vitreous endosperm formation. Illumina exon- and RNA-seq revealed a 1.2 Mb deletion encompassing the 27- and 50-kD γ-zein genes on chromosome 7, and a deletion of at least 232 Kb on chromosome 9. Protein body number was reduced by over 90% while protein body size is similar to wild type. Kernels hemizygous for the γ-zein deletion had intermediate 27- and 50-kD γ-zein levels and were semi-vitreous, indicating haploinsufficiency of these gene products in *opaque-2* endosperm modification. We showed that such haploinsufficiency for 27-kD γ-zein function does not apply to the modification in B73 wild type vitreous endosperm formation. The γ-zein deletion further increased lysine in QPM in its homozygous and hemizygous states. This work identifies 27-kD γ-zein as an *opaque-2* modifier gene within the largest QPM QTL, and may suggest the 50-kD γ-zein also contributes to this QTL. It further demonstrates that genome-wide deletions in non-reference maize lines can be identified through a combination of assembly of Illumina reads against the B73 genome and integration of RNA-seq data.

INTRODUCTION

Maize opaque endosperm mutants have been studied because of the nutritional improvement that they often show, as well as for the insight they provide into the kernel

hardening; an essential quality trait for most of the grain's uses. The most well-known opaque mutant, *opaque2* (*o2*), has been widely studied because of its increased lysine and tryptophan level (Mertz et al., 1964) associated with the reduced synthesis of alpha zeins. The *O2* gene encodes a transcription factor that regulates α-zein gene expression (Schmidt et al., 1990), although all zein proteins are indirectly reduced. O2 also regulates pyruvate Pi dikinase (Maddaloni et al., 1996) and many other genes (Jia et al., 2013). Although the soft kernels and yield penalty of *o2* inhibited its commercial success, subsequent breeding projects, including those in Mexico (Vasal et al., 1980) and in South Africa (Geevers and Lake, 1992), led to the development of hard kernel *o2* varieties called Quality Protein Maize (QPM). QPM kernels generally have low levels of α-zeins and high levels of lysine and tryptophan, but the genetic basis of *o2* endosperm modification is poorly understood.

One prominent feature of QPM endosperm is accumulation of the 27-kD γ-zein at two- to threefold higher levels than in wild type and *o2* (Geetha et al., 1991; Lopes and Larkins, 1991; Wallace et al., 1990). Although the genetic and biochemical mechanisms responsible for this increase are unknown, the degree of endosperm vitreousness in QPM correlates with the level of 27-kD γ-zein protein (Lopes and Larkins, 1991). Furthermore, the 27-kD γ-zein gene, along with the closely linked 50-kD γ-zein, maps to the most significant QTL for endosperm modification located on chromosome 7 (Holding et al., 2008; Holding et al., 2011; Lopes and Larkins, 1995; Lopes et al., 1995). QPM endosperm accumulates larger numbers of small, γ-zein rich protein bodies. This is proposed to allow the formation of a rigid glassy matrix similar in texture to mature wild

type endosperm. RNAi reduction of both 27- and 16-kD γ-zeins in QPM caused opaque reversion, and thus supported the suggestion that 27-kD γ-zein contributes to endosperm modification (Wu et al., 2010). However, the extent to which 27-kD γ-zein is alone sufficient as a modifier is unknown. Furthermore, the possible role of the 50-kD γ-zein in modification has not been previously addressed. The molecular characterization of opaque mutants has shown that vitreous endosperm formation depends on more than just correct abundance and spatial organization of zein proteins (Holding and Larkins, 2006; Holding et al., 2010; Holding et al., 2007; Myers et al., 2011; Wang et al., 2012). Furthermore, mapping studies indicate there are several QTLs for endosperm modification in QPM (Holding et al., 2008; Holding et al., 2011; Lopes and Larkins, 1995; Lopes et al., 1995).

We used γ-irradiation of K0326Y QPM as a means to identify genes corresponding to *o2* modifiers and genes generally required for vitreous endosperm formation. For identification of DNA sequence deletions in the induced mutants, we combined exon-capture DNA sequencing and RNA-seq. Among the opaque variants induced in QPM, one had a deletion that eliminates both the 27- and 50-kD γ-zein loci which are within a significant QPM QTL on chromosome 7 (Holding et al., 2008; Holding et al., 2011). By demonstrating the dosage dependent action of 27-kD γ-zein in *opaque-2* endosperm modification, this mutant highlights the potential of this approach for identifying other *o2* modifier genes as well as genes involved in a variety of aspects of endosperm development.

METHODS

Radiation dosage testing and γ-radiation mutagenesis

K0326Y QPM seed were equilibrated to 13% moisture as previously described (Hossain et al., 2004) and treated in batches of approximately 50 seeds with various dosages of γ radiation (15 Gy to 35 Gy). After irradiation the seeds were immediately planted in the greenhouse in potting soil. Plant survival was determined as the proportion of seedlings compared to non-irradiated control seeds one month after sowing. Radiation dosage was set at 25 Gy based on a 75% survival rate and was used for full-scale mutagenesis of 1909 seeds. These seeds were planted in the field at the University of Nebraska East Campus farm in summer 2010.

A population of 1909 K0326Y QPM seeds, treated with 25 Grays (Gy) of γ-radiation, were propagated in the field in 2010 and 1108 germinated (58%). The radiation dosage was previously optimized to produce 75% survival rate under controlled greenhouse condition using a small-scale (~50 seeds per dosage) experiment.

Screening of M2 and heritability testing in M3

Adequately filled M2 ears (more than approximately 30 kernels) that did not have fungal contamination were screened for dominant kernel phenotypes before shelling and after shelling on light box. Zein profiles were obtained from putative opaque mutant kernels to prioritize lines for more urgent greenhouse heritability testing. Twenty kernels from all M2 ears were propagated in the field in summer 2011. This enabled heritability testing of dominant M2 phenotypes, rescue of M2 ears with few kernels and identification of

recessive mutants appearing in the M3 generation.

Genotyping, RT-PCR, zein protein analysis, and fixation for microscopy

Plants or seeds were selected for the γ-zein deletion using leaf or embryo genomic PCR using DNA extracted according to the urea method (Holding et al., 2008). For RT-PCR RNA was extracted and DNase treated cDNA synthesized as previously described (Holding et al., 2011). All primers used for genotyping and RT-PCR are not shown. Zeins were extracted from mature and developing endosperm as previously described (Wallace et al., 1990). Endosperms of kernels genotyped for the deletion were processed as previously described (Guo et al., 2013). For the 18 DAP stage, kernels were removed, without damaging the base of the kernels, using a razor blade and treated as follows. Kernels were placed embryo side down on a clean glass plate and sliced in half longitudinally through the embryo. The embryo halves were placed immediately into DNA extraction buffer on ice, one half of the endosperm was frozen in liquid nitrogen for protein analysis and the other half kernel was dissected for fixation. Keeping the pericarp intact, the half kernel was placed cut side down and a 1-2 mm longitudinal central section was taken and placed into 5mls fixative (2% sucrose, 0.1M sodium cacodylate pH 7.4, 5% glutaraldehyde). Samples were fixed at 4°C for at least one week while genotyping and SDSPAGE of zeins were conducted to identify transgenic kernels for further processing. Embryo DNA was extracted using microcentrifuge scale urea DNA extraction.

Immunogold TEM and protein body counts

Selected kernel segments were further dissected following fixation to remove the apical and basal kernel portions leaving two, 1mm mid-kernel pieces with pericarp still attached for orientation. Samples were infiltrated in LR White resin (EMS) before sectioning. 90nm sections were cut using a Diatome diamond knife and a LKB ULTROTOME III microtome and attached to carbon and Formvar coated grids (EMS). For immunogold labeling, grids were hybridized in 50 µl drops on Parafilm strips as follows. The grids were incubated for 30 min in blocking solution, pH 8.2, containing 0.2% BSA and 0.06% Tween20 in 20 mM Tris-HCl and 500 mM NaCl. Rabbit primary antisera were added (α-zein at 1/2000 and γ-zein at 1/100) and incubated overnight at 4°C before three 10 min washes in 100 µl blocking solution. Secondary antibody (goat anti-rabbit 15 nm gold conjugated from EMS) was added at 1/20 dilution for 1 hour at room temperature. The grids were washed three times and rinsed three times in deionized H2O and air (Holding et al., 2008) viewed on a Hitachi 7500 TEM.

Protein body images were standardized by viewing the 4th sub-aleurone starchy endosperm cell layer. At the 18-DAP kernel stage, 1st starchy endosperm cell layer had few protein bodies and increased in size and number in the layers internal to this. Estimations of protein body number were standardized by making counts within 100 µm square regions within 4th sub-aleurone cell layer cells away from the cell wall and nucleus.

Exon-seq

Genomic DNA was extracted from leaf tissue using a modified urea extraction protocol

(Holding et al., 2008). DNA concentration and purity was assayed with nanodrop and 100µg crude DNA purified using Qiagen DNA clean Kit. 10 µg samples of purified DNA was used for exon capture. DNA library construction, Exome capture and QC validation were carried out as a paid service by the Genomic Technologies Facility at Iowa State University. For DNA library construction 3 µg DNA samples were sheared with nebulizers and prepared incorporating bar codes according to Illumina TruSeq DNA Sample Prep kit's protocol (Cat# FC-121-2001). For quality control of the pre-capture libraries, a Bioanalyzer 2100 was used to establish library size and quantity. For hybridization each library was diluted to 30 ng/µl. Fragmented DNA from four opaque deletion mutants and the progenitor K0326Y QPM parent were used to make Illumina true-seq DNA libraries, which allowed incorporation of bar-codes within the adaptors and pooling of the five samples. The five libraries (300ng per library) were combined together for one array hybridization using the Nimblegen Zm B73 HX3 exome capture array (1.5µg total per array). The procedure of hybridization followed NimbleGen Array User's Guide (Plant Sequence Capture Illumina Optimized protocol, v 1, 2010). The eluted libraries were subjected to 10 cycles of amplification using the TruSeq enrichment PCR kit. For validation of hybridization efficacy, the ratios of five random genes were tested by real time q-RT-PCR in the pre- and post-captured libraries as part of the service at the ISU genomics facility (not shown). Libraries were then sequenced (100bp single end reads) using the Illumina Genome Analyzer 2 service at the University of Nebraska, Lincoln genomics core facility. Short reads were discarded if their average quality scores are smaller than 30. The adapter sequence on 5'-end and six nucleotides on the 5'-end of 7-base bar codes were used to determine the trimming position, and deconvolute the

reads. After trimming the adapter sequence and bar codes, the grouped short reads were mapped against the B73 genome (ZmB73_RefGen_v2 from maizesequence.org) using Bowtie (Langmead et al., 2009), allowing up to two base mismatches per read. Reads mapped to multiple locations were discarded. Numbers of reads in exons were counted using the HTSeqcount tool using B73 gene annotations (ZmB73_5b_FGS from maizesequence.org) and the 'union' resolution mode was used. For pair-wise comparisons, the edgeR package (Robinson et al., 2010) with TMM normalization method (Robinson and Oshlack, 2010) was used to analyze the numbers of reads aligned to exons.

The use of a Nimblegen exome array to purge repetitive and intron sequences and specifically select for exon sequences, resulted in the template size of each mutant being reduced from 3.2 Gb to ~50 Mb (64-fold simplification) and allowed us to sequence this pool in a single Genome Analyzer II flow cell lane, which has a coverage of 3.6 Gb. After deconvolution of reads from the five lines, they were assembled against the B73 genome. The non-isogenic nature the QPM background compared with B73 precluded the possibility to assemble reads and infer sub-exon mutations in this population, but we were able to effectively score for presence/absence of exons. Deletions were called according to an algorithm that considered both the number of reads for a given exon in the non-mutagenized QPM and the ratio of reads between non-mutagenized QPM and a mutagenized line. An exon was automatically called a deletion candidate if its P-value was smaller than 10^{-15} (30% of candidates fit this criterion). The other 70% of called candidates had P-values $\in$ 10^{-15} to 10^{-5}. These candidates were only called when the

number of reads for the mutagenized line were not larger than 0, 1, 2, 3, and 5 when the numbers of reads for the non-mutagenized reference line were in ranges of (30-50), (50-70), (70-100), (100-150), (150 and above) respectively. For example, a candidate with 125 reads in non-mutagenized line must have three or less reads in the mutagenized line.

RNA-seq

Even though we are mostly interested in the loss of endosperm specific transcripts, total RNA was extracted as previously described (Holding et al., 2011) from leaf, root and whole seed. RNA was purified using Qiagen RNeasy cleanup kits and a pool was made for RNA-seq in order to identify transcripts abolished from multiple tissues. 10 μg of RNA was used for preparation of Illumina cDNA libraries and 100bp reads were obtained from Illumina Genome Analyzer 2 platform using RNA from non-mutagenized K0326Y, line 107 and one other deletion mutant per flow cell lane. For the RNA-seq reads, the same method as exon reads was used to deconvolute the bar codes, trim the adapter sequences, map to the reference genome, and analyze the abundance of reads in each gene. A threshold value for fold-change differential expression was set at $\log2 > 1$ (two-fold actual values) and adjusted *P*-values < 0.001 for the null hypothesis although we show genes changes at least $\log2 > 2$ (four-fold actual values). Differentially expressed genes were annotated using the Gene Ontology database.

Sequence data from this chapter can be found in the GenBank/EMBL data libraries under accession numbers AF371263, DQ400403, AF371261, and NM001174129 for primer design in the 50-, 16-, and 27-kD gamma-zein genes and the tublin control gene,

respectively. All other genes are listed by their maize genome descriptors (eg. GRMZM2G398628).

RESULTS

Identification of a 27- and 50-kD γ-zein protein null among novel opaque deletion mutants

We recovered 305 M2 ears of variable seed fill of which 293 were successfully propagated to M3 families (~20 ears each). Although single M2 ears were potentially affected by non-genetic environmental factors, screens of adequately filled M2 ears were conducted for dominant seed mutant phenotypes and identified more than 30 lines having a substantial proportion of opaque, small, defective, empty pericarp or rough kernels. 27 lines exhibited clear segregation for opaque mutants. Of these 27 lines, opaque kernels in 17 lines were not viable and presumed to be pleiotropic mutations. The other ten lines had fully viable opaque kernels and were thus candidates for factors specifically involved in *o2* endosperm modification and/or vitreous endosperm formation in general. Three of these ten confirmed mutants, including line 107 (Figure. 1), 198 (Figure. 1, chapter 3) and line 12, exhibited mutant kernels in the M2 ears while the others were first observed in the M3 and resulted from M2 ears which either appeared normal or had too few kernels to screen. To identify segregating kernel mutation, the ear must be healthy and well filled to exclude non-genetic environmental effects on kernel development. Analysis of zein proteins revealed several different phenotypic classes affecting zein composition. These included no apparent change in zeins (line 12, 121 and 134), generalized decreased abundance of all zeins (lines 112, 198 and 224), and loss of residual 22-kD α-zeins that

present in K0326Y QPM (lines 8, 44, 133 and 224) (Figure. 2). Line 107 exhibited a striking absence of the 27- and 50-kD γ-zeins (Figure. 3A), which was confirmed with immuno-blot analysis (Figure. 3B). Given the suggested role of 27-kD γ-zein in endosperm modification in QPM, we focused on the characterization of this mutant. As expected, 27-kD γ-zein was significantly increased in K0326Y QPM compared with W64A WT (Figure. 3A). Interestingly, although 50-kD γ-zein is not as abundant as the 27-kD γ-zein protein and does not appear to be involved in protein body initiation (Guo et al., 2013), similar to 27-kD γ-zein, we consistently observed the protein to be increased in QPM compared with wild type (Figure. 3A and 3B).

Complete lack of 27-kD γ-zein in line 107 dramatically reduces protein body initiation but further improves protein quality

To investigate the effects of the γ-zein deletion on protein body initiation, morphology and zein distribution, we conducted immunogold TEM of 18 DAP developing endosperm. Low magnification (1000x) of the W64A wild type control showed that protein bodies are densely clustered by the 4th sub-aleurone cell layer (Figure. 4A). Thus, for quantitative counts and qualitative analysis protein body morphology and composition, we focused on the 4th subaleurone cell layer. Starch grains in this layer remained smaller and less abundant than in more central endosperm regions which would have impaired the ability to quantify protein body number. Cell layers internal to the 4th layer could often not be reliably deciphered because of the oblique and irregular cell wall planes as well as the relatively lower contrast of the immunogold TEM compared with standard TEM. Non-mutagenized QPM accumulated smaller protein bodies 30-40% of

the W64A WT size but in higher number than WT (Figure. 4B, Figure. 5D-F, Table 1), which is consistent with the suggested role of increased 27-kD γ-zein in QPM (Geetha et al., 1991). Protein bodies in QPM are darker staining due to deposition of γ-zein throughout the protein body (Figure. 5E) where it co-localizes with α-zeins (Figure. 5F). Elimination of 27- and 50-kD γ-zein proteins in line 107 resulted in almost undetectable numbers of protein bodies at low magnification (Figure. 4C) due in part to their lack of contrast compared with WT or QPM protein bodies (Figure. 5G and H). At high magnification, and with the aid of α-zein labeling (Figure. 5I), line 107 protein bodies were identified as being similar in size to WT and with irregular, undulating surfaces. Protein body number in line 107 is dramatically reduced compared with WT and QPM Figure. 4C, Table 1). This is similar to results obtained by RNAi suppression of 27-kD γ-zein synthesis (Guo et al., 2013), except that suppression of protein body initiation is even more severe in the complete absence of 27-kD γ-zein (Figure. 5G-I). Consistent with RNAi results (Guo et al., 2013), this shows that 16-kD γ-zein is not sufficient for supporting the normal level of protein body initiation. The increased protein body size in line 107, relative to its QPM progenitor (Figure. 5G-I), is likely due to the need to package substantial 19-kD α-zein in a smaller number of protein bodies. The irregular undulating protein body shape (Figure. 5G-I) may be due to the improper packaging of the hydrophobic 19-kD α-zein. The γ-zein antibody, which was raised against the 27-kD γ-zein, does not detect γ-zein in line 107 protein bodies that explains the lack of protein body contrast.

We measured lysine and tryptophan concentrations in mature kernels of the 107 mutant

and compared it with the K0326Y QPM progenitor, W64A wild type and W64A*o2*. Comparison of W64A WT and *o2* kernels showed the expected levels of *o2* improvement for these amino acids (Table 2). Improvement of the lysine and tryptophan percentage in QPM was not as pronounced as W64A*o2* probably as a result of higher level accumulation of lysine-devoid 27-kD γ-zein. However, when 27-kD γ-zein was absent, the concentration of these amino acids was similar to W64A*o2*. Consistently, the partially modified semi-opaque 107 kernels had a higher level of lysine than QPM, suggesting that partial accumulation γ-zein can present a breeding strategy to balance protein quality and endosperm texture.

Line 107 shows absence of 27- and 50-kD γ-zein genes and haploinsufficiency for *o2* endosperm modification

We sought to determine if the absence of 27- and 50-kD γ-zein proteins resulted from an indirect effect on zein synthesis in general, a regulatory defect or the physical loss of the genes. RT-PCR showed that transcripts of both genes were undetectable in endosperm RNA of line 107 (Figure. 6A). Levels of the unlinked 16-kD γ-zein were normal ruling out the possibility that zein gene expression was generally affected. Genomic PCR did not detect the 27- or 50-kD zein genes while the 16-kD γ-zein gene was present (Figure. 6B) showing that the protein absence was not a regulatory effect but due rather to the loss of these genes. Given that the 27- and 50-kD γ-zein genes are closely linked (being separated by approximately 27 Kb (Holding et al., 2008)) the most likely scenario was that a deletion covered both genes, which was confirmed by sequencing exon captured genomic DNA (below).

While most of the heritable opaque mutants we recovered appeared to segregate according to Mendelian ratios in the M3, line 107 had more than 25% opaque kernels and obvious variability in the level of opacity (Figure. 1B). With careful examination on a light box, the number of fully opaque kernels (Figure. 7) was found to be approximately 25% (46/190, 9/38 and 14/56 in three M3 segregating ears). Kernels in this class were homozygous for the deletion mutation since they gave rise to 100% opaque kernels when propagated (Figure. 1C). Another class of kernels, making up less than 50%, were opaque with a vitreous cap (Figure. 7). These were determined to be hemizygous for the γ-zein deletion, since they gave rise to segregating ears when propagated (Figure. 1B). The remainder of the expected 50% hemizygous kernels from these ears comprised less severe semi-opaques, which could not be reliably scored. SDS-PAGE analysis of zein proteins from vitreous, semi-opaque and fully opaque kernels from a segregating ear showed that partial opacity results from reduced accumulation of the 27- and 50-kD γ-zein proteins (Figure. 7). This suggests that expression from a single allele of these two genes is haploinsufficient for full *o2* endosperm modification. Importantly, unlike the majority of mutants which first segregated in the M3 ears, the M2 ear for line 107 segregated 1:1 for normal and opaque with vitreous cap (hemizygous) kernels which supports the haploinsufficiency explanation and is consistent with the ontology of the mature embryonic shoot apical meristem (Johri and Coe, 1983; Poethig et al., 1986) (see discussion).

Haploinsufficient role of γ-zein does not apply to vitreous endosperm formation in

wild type endosperm

By crossing line 107 to B73 wild type, we sought to address whether this haploinsufficient role of γ-zein also applies to the vitreous endosperm formation in wild type endosperm. The *o2* mutation and the γ-zein deletion segregated independently in F2 ears from the self-pollinated F1 plant and since *o2* modifier genes are expected to segregate independently, we expected *o2* to be unmodified or partially modified. Analyzing zein proteins from opaque and vitreous kernels of F3 ears showed that the *O2* genotype as well as the γ-zein deletion phenotype could be related to the kernel phenotype (Figure. 8). Since 22-kD α-zein and 15-kD β-zein are both transcriptional targets of O2, their low levels readily distinguish homozygous *o2* (lane 8 or 9, for example). Opaque kernels in a typical F3 segregating ear have the genotype of homozygous or heterozygous O2 with homozygous γ-zein deletion (lane 7), homozygous mutant for both *o2* and γ-zein deletion (lane 8) or homozygous *o2* mutant with homozygous γ-zein non-deletion (lane 9). Semi-vitreous kernels were not observed as they were in line 107. By surveying multiple fully vitreous kernels, some were homozygous wild type for O2, but apparently hemizygous for the γ-zein deletion since they had an intermediate amount of 27-kD γ-zein (lane 4). But these kernels were fully vitreous, not semi-vitreous as is the case for hemizygous γ-zein deletion kernels in the QPM background. In all, this study suggests that in B73 wild type line, one γ-zein copy is sufficient for the vitreous endosperm formation and the haploinsufficient role of 27-kD γ-zein only applies to *o2* endosperm modification in QPM where maximal 27-kD γ-zein is needed.

Exon- and RNA-seq identifies a deletion spanning the 27- and 50-kD γ-zein genes

In order to confirm that the 27-kD and 50-kD γ-zeins were within a single deletion and define the size of that deletion, we used Illumina sequencing of exon-enriched genomic DNA. Deletions were called according to an algorithm that considered both the number of reads for a given exon in the non-mutagenized QPM and the ratio of reads between non-mutagenized QPM and a mutagenized line (see methods). The relatively non-stringent criteria returned deletion candidates on all chromosomes apart from chromosome 10 (data not shown). We tested all of the candidates using genomic PCR and RT-PCR (data not shown) and established that all except those genes within the chromosome 7 and 9 deletions described below were false positives. As further confirmation, none of the genes identified as false positives were identified as differentially expressed using RNA-seq (data not shown).

By combining exon read data with genomic and RT-PCR we identified two deletions in line 107. One of these is in chromosome 7, Bin 7.02 and is approximately 1,236,000 bp (Figure. 9). Using genomic PCR on exon deletion candidates, we showed that there are at least 11 predicted genes in this region that are absent in line 107. RT-PCR revealed that at least seven of these genes are expressed in normal developing endosperm but not detectable in RNA of line 107 (Figure. 9). Two genes close to the 5' end of the deletion encode the 27- and 50-kD γ-zeins. Four of the 11 genes are either endosperm non-expressed, expressed at a low level or are pseudogenes. Four other genes were confirmed as deleted in this region by comparing exon- and RNA-seq reads (data not shown). We used primers within flanking genes (Figure. 9) as well as intergenic primers to estimate

the extremities of this deletion. A smaller deletion of at least 231,968 bp was identified on chromosome 9, Bin 9.03 (Figure. 10). This region contains five genes, four of which have expression abolished based on RT-PCR. Expression of the fifth gene was not detectable in the non-mutagenized control. Genomic and RT-PCR tests of other exon deletion candidates (data not shown) identified elsewhere in the genome by low or absent read number did not confirm additional single or multiple exon deletions.

The RNA-seq comparison of K0326Y QPM and line 107 had three purposes: confirmation of deletions identified with exon-seq; identification of additional deletions that exon-seq may have missed; and potential identification of indirect transcriptional effects of putative gene deletions, as a consequence of defective transcription factors. For this, we mixed RNA in equal proportions extracted from whole developing seeds (endosperm and embryo), juvenile leaf and juvenile root. This potentially allowed us to identify transcript alterations resulting in sporophytic tissues as well as seed tissues. Since our goal was to identify complete losses or substantial changes in expression, rather than minor fold-changes, a single RNA-seq library was prepared from K0326Y QPM and line 107, and this allowed bar-coding and pooling with other deletion mutants. The genes with transcripts that were decreased or increased in line 107 based on RNAseq, along with their exon-seq reads are not shown. Twenty genes were increased in abundance more than four-fold in line 107 compared with K0326Y QPM whereas 70 genes had transcripts more than four-fold reduced. Within these 70 were genes shown to be within the chromosomes 7 and 9 deletions (data not shown). Of the five genes within the chromosome 9 deletion, only one gene (GRMZM2G099648; a ubiquitin ligase) had

RNA-seq reads. Of the other differentially expressed genes with chromosome coordinates not within the chromosome 7 and 9 deletions, none qualified as candidate deletions. However, several differentially expressed genes identified with RNA-seq were not detected with exon-seq in either K0326Y or line 107. We tested all of these genes with at least two pairs of genomic PCR primers each and showed that they were not gene deletions missed during exon-seq (data not shown).

Zein genes are expressed at high levels and this was reflected in the reads measured for the 27-kD and 50-kD γ-zein genes in K0326Y QPM (386,441 reads and 180,567 reads respectively). As expected, reads for these genes in line 107 were drastically reduced to 0.2% and 0.1% of control respectively, but surprisingly, they were not completely absent as might be expected for deletion of single copy genes (869 and 284 reads respectively). These genes share extensive identity with the 16-kD γ-zein, but at a lower level than our allowable threshold for RNA-seq reads (98bp perfect match of the 100bp read length). However, some nucleotides within the γ-zein reads had sub-optimal scores and if some had been incorrectly called for 16-kD γ-zein reads, this could give rise to reads incorrectly annotated as 50-kD or 27-kD reads. Alignment of K0326Y sequences to the B73 reference genome is also a possible source of errors. In addition, since K0326Y and line 107 were sequenced in the same pool in the same flow cell lane, these reads could, in theory, be the result very low level incorrect parsing of bar code sequences.

Three alpha zein genes were around six to 16-fold down-regulated in line 107 (data not shown). One of these is annotated as a Z1A (19-kD) family member and two are Z1C

(22-kD) family members. Z1A genes are not regulated by O2 and the Z1C genes identified here are also likely O2-independent. The basis for this conclusion is that their expression in K0326Y (modified o2/o2) is still substantial and the O2 transcript in both K0326Y and line 107 was undetectable by RT-PCR and qRT-PCR (data not shown). Consequently, these alpha zeins may be repressed though an unknown feedback mechanism which results from reduced levels of γ-zein synthesis. Such a mechanism was also proposed as a result of the recent observations of indirect suppression α-zein gene expression caused by various γ-zein RNAi constructs (Guo et al., 2013).

DISCUSSION

The 27-kD γ-zein acts in a dosage-dependent manner to modify *o2* endosperm

The feasibility of radiation mutagenesis to identify *o2* modifier genes was demonstrated since one line of a population of 293 M2 families, had a deletion of a suspected modifier gene, the 27-kD γ-zein as well the nearby 50-kD γ-zein, and caused complete opaque reversion in QPM. Null mutant alleles have not been previously reported for either of these genes, although knock-down lines have been generated with RNAi by ourselves and others (Guo et al., 2013; Wu and Messing, 2010). RNAi studies have allowed dissection of the overlapping and distinct roles of the different γ-zeins but their interpretation is not simple, because the transgenic lines are not null. However, their phenotypes have suggested that the 27-kD γ-zein has a specialized role in protein body initiation (Guo et al., 2013). Crossing an RNAi event which reduces both the 27- and 16-kD γ-zeins to K0326Y QPM, caused complete opaque kernel reversion, and thus is consistent with abundant previous data suggesting that 27-kD γ-zein plays a key role in

endosperm modification (Geetha et al., 1991; Lopes and Larkins, 1991; Lopes and Larkins, 1995; Lopes et al., 1995; Wu et al., 2010).

The 27- and 50-kD γ-zein null line reported here provides more definitive proof of essentiality of the 27-kD γ-zein for QPM development. It also invokes the possibility that the 50-kD γ-zein is involved in endosperm modification. Despite the relatively lower abundance of the 50-kD γ-zein, it is readily detectable in wild type endosperm and becomes almost undetectable in *o2* endosperm (Figure. 3). Furthermore, during endosperm modification in QPM, this protein is not only restored but is increased relative to wild type. Like that of increased 27-kD γ-zein transcript and protein (Holding et al., 2011), the mechanism causing this increase is unknown. It is possible that pre- or post-transcriptional regulation could affect these genes independently, or that they could be co-regulated. In either case, these genes are so closely linked that it is not currently possible to separate their contributions to the QTL on chromosome 7.02.

The 27-kD γ-zein deletion also extends our knowledge of its role in protein body initiation. While protein body number is reduced to about 40% of the wild type number by 27-kD γ-zein RNAi (Guo et al., 2013), the knock-out mutant reduced the protein body number to about 12% of wild type and about 7% of QPM protein body number. Vitreous endosperm formation in modified *o2* is thought to be driven, at least in part, by the initiation of a larger number of protein bodies as a result of increased 27-kD γ-zein accumulation (Dannenhoffer et al., 1995; Gibbon and Larkins, 2005). These protein bodies remain small with γ-zein located throughout (Figure. 5), and have an even round

shape due to their appropriate packaging of residual α-zein (Figure. 5). However, in the absence of the 27-kD and 50-kD γ-zeins, the protein body size is similar to wild type with highly irregular shape. The size increase likely results from the packaging of residual α-zein in less than 10% of the number of protein bodies (Figure. 5) and the distortions likely result from hydrophobic α-zeins, especially the abundant 19-kD α-zeins, not being correctly packaged by the remaining γ-zeins. Protein body shape distortions also result from dominant negative α- and γ-zein mutants which interfere with correct zein trafficking into the lumen of the ER (Coleman et al., 1997; Kim et al., 2006; Kim et al., 2004). Similarly, low-level accumulation of an unprocessed α-kafirin protein causes protein body reticulation in the *high digestibility high lysine* sorghum mutant (Wu et al., 2013).

Unlike the highly duplicated α-zein genes, whose expression collectively results in massive levels of α-zein protein, the γ-zeins are encoded by single copy genes and in the case of 27-kD γ- zein, protein levels approach those of α-zeins. What drives this high level expression as well as it's even greater expression in QPM is not known. The degree to which γ-zein protein levels are limited by pre- and post-transcriptional factors is also not known. Our results show that one copy of the 27-kD γ-zein gene is not sufficient for full modification, since hemizygous seeds accumulate intermediate amounts of the protein and are only partially modified. This haploinsufficiency suggests that maximal transcription from both 27-kD γ-zein gene copies is required to create a fully vitreous endosperm. However, the synthesis of 27-kD γ-zein is also likely limited by amino acid availability and translational capacity since 27-kD γ-zein over- expression constructs do

not further increase the protein (D. Holding, unpublished data). The γ- zein deletion is currently being introgressed into B73 wild type to separate it from the effects of the *o2* mutation and the modifier genes. This will provide insight into whether this same critical threshold for γ-zein abundance also exists for normal vitreous endosperm formation as it does for endosperm modification in QPM.

The increased lysine and tryptophan concentrations of the QPM γ-zein deletion emphasize one of the drawbacks of QPM in relation to un-modified *opaque-2*. Since 27-kD γ-zein is devoid of lysine, its increased level in QPM reduces the nutritional quality compared with starchy unmodified *o2*. Our results show that this particular QPM inbred has lysine and tryptophan levels intermediate between W64A*o2* and wild type. However, hemizygous γ-zein deletion kernels had similar amounts of these amino acids to unmodified *o2* while exhibiting greater levels of vitreousness. This may suggest that selection of lines with partially accumulating γ-zein may be a novel strategy for breeding maize lines with both increased lysine and an acceptably vitreous kernel texture.

The gamma-zein deletion line demonstrates the potential for identifying other *o2* modifiers as well as for general maize seed functional genomics

We used gamma irradiation mutagenesis to study genes affecting vitreous endosperm formation (modification) in the hard kernel *opaque-2* variant, K0236Y QPM. In addition to *o2*-modifier genes, this approach potentially allows functional genomic dissection of genes more generally involved in endosperm formation as well as seed and plant development. In a population of 293 M3 families, we identified 27 mutants segregating

for opaque kernel phenotypes. Although M2 screens did identify at least three dominant opaque kernel mutants, which were true breeding from the M2 to the M3 and M4, the majority of opaque mutants were first observed segregating in the M3 ears. The likely reason for this is that different populations of cells within the mature maize embryo shoot apical meristem give rise to the ear and tassel (Johri and Coe, 1983; Poethig et al., 1986). A hemizygous DNA deletion in an ear progenitor cell within the shoot apical meristem of a mature embryo will ultimately result in a mixture of hemizygous and wild type kernels in an M2 ear because the tassel progenitor cells giving rise to the pollen would not have the mutation. Thus, only dominant mutations would be visible in the M2 ears. The number of tassel and ear progenitor cells in the meristem was given as four ($\pm$1) for the tassel and one to three cells for the ear (Johri and Coe, 1983). These numbers would be expected to produce variable phenotypic ratios for dominant mutants in M2 ears. Line 107 and line 12 (not shown) had apparent 1:1 ratios of normal to mutant seeds (19:19 and 164:174 respectively) and may suggest that the mutation arose in one meristem initial cell which gave rise to the whole ear. The other dominant mutant (line 198) had a lower ratio of M2 mutant kernels (230:52) perhaps suggesting that the mutation arose in one of several ear or tassel initial cells.

Using mutants that had viable kernels taken from segregating M3 ears, and gave rise to 100% mutant M4 ears, we were able to assume homozygosity and directly assay for gene loss using Illumina sequencing of genomic DNA. To simplify the sequencing template, we enriched for the exon fraction, using exon capture hybridization to a Nimblegen B73 maize Exome array. This simplification of sequencing template enabled pooling of five

mutants and sequencing in a single Genome Analyzer II flow cell lane. Despite this sub-optimal Illumina platform (compared with the Hi-Seq 2000 platform) and assembling K0326Y QPM sequences against the non-isogenic B73 genome, we were able to call and test deletions ranging from single to multiple genes. RNAseq allowed verification of the deleted genes and identified genes with expression that was indirectly affected by DNA deletions.

Prospects and limitations for maize seed functional genomics using radiation mutagenesis

The data presented here strengthen genetic, biochemical and transgenic evidence suggesting that the 27-kD γ-zein, possibly acting with the 50-kD γ-zein, is an *opaque-2* modifier gene. As with transposon or EMS mutagenesis, in which multiple genes are mutated, deletion mutagenesis requires reverse genetics to formally link mutations to phenotype. Typically, this involves isolation of independent mutant alleles in the gene suspected of causing the phenotype or complementation or knock-out using transgenic constructs. The knock-out mutants of the 27- and 50-kD γ-zeins described herein are the only known mutants in these genes. However, we showed the essentiality of 27-kD γ-zein in modification of K0326Y QPM endosperm using RNAi (Wu et al., 2010). Furthermore, RNAi lines which dramatically reduce the level of 27-kD γ-zein support the role of this protein in vitreous endosperm formation in a normal vitreous endosperm formation (Guo et al., 2013; Wu and Messing, 2010).

In addition to identifying genes involved in vitreous endosperm formation, characterizing

more opaque mutants in this QPM population will allow appraisal of the heterogeneity and frequency of radiation induced deletions and their utility for maize functional genomics when combined with current reverse genetics resources and ever improving sequencing capabilities. In addition to large deletions, radiation can also cause sub-gene scale mutations (Naito et al., 2005), but we could not detect these without a K0326Y QPM reference genome. A second, much larger, deletion population has been created that has 1784 mutagenized families in the B73 reference genetic background. Many apparent viable opaque and small kernel seed mutants are currently being tested for heritability. Because of the isogenicity of these mutants with the B73 reference genome, we will be able to detect sub-gene mutations in this population, and exon- and RNA-seq reads can be assembled and compared for identification and appraisal of the power of deletion mutagenesis for creating and utilizing mutations ranging from a few base pairs to mega-base pairs.

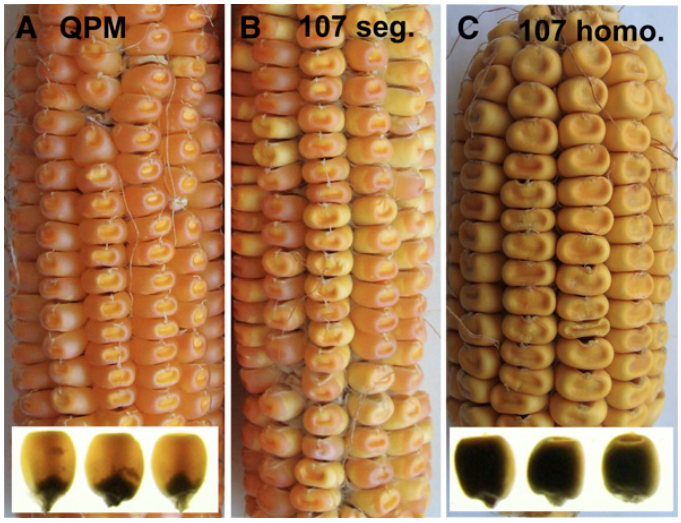
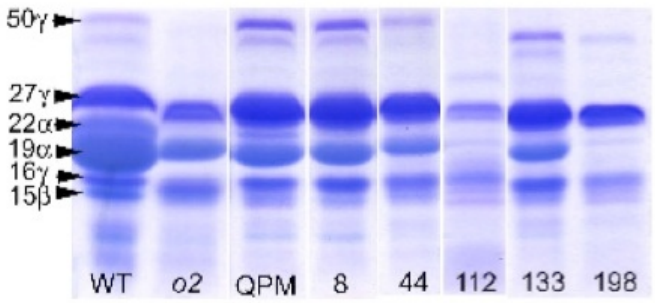


Figure 1. M3 ear phenotypes of K0326Y QPM control ear and segregating and homozygous K0326Y deletion line 107. Inserts in A and C show light box phenotypes.



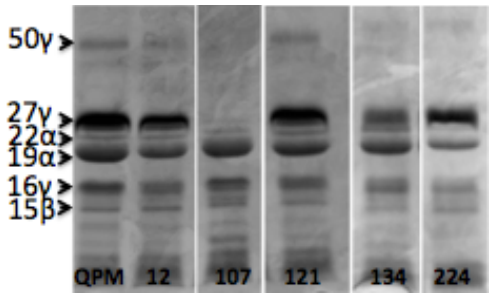


Figure 2. zein profiles of mutants compared with W64A wild type, W64A o2 and K0326Y QPM.

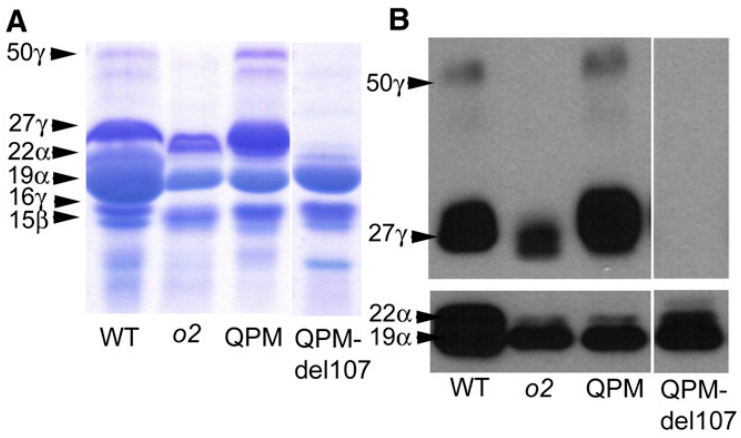


Figure 3. A, SDS-PAGE analysis of zein proteins from mature kernel of deletion line 107 compared with K0326Y QPM and the W64A wild type and W64Ao2. B, Western analysis of zein proteins loaded at1:1,000 dilution of gel shown in A, probed with total α-zein antiserum (1:10,000) and anti-27-kD γ-zein antiserum (1:2,000). WT,Wild type.

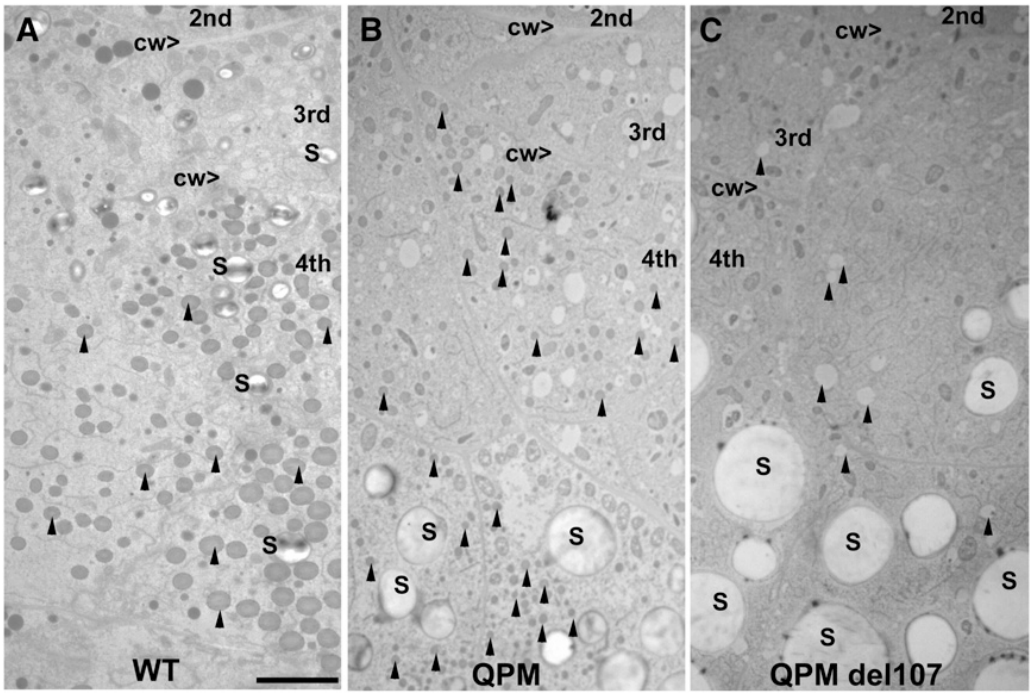


Figure 4. Low-magnification TEM images from immunogold labeling showing distribution of protein bodies in representative 18-DAP endosperm of the wild type, K0326Y QPM, and QPM deletion line 107. Subaleurone cell layers are labeled with 2nd, 3rd, and 4th. Cell walls are labeled with cw., starch grains are labeled with S, and selected protein bodies are labeled with vertical arrows. Depicted protein bodies in C show connections to the endoplasmic reticulum. Bar = 5 mm. WT, Wild type.

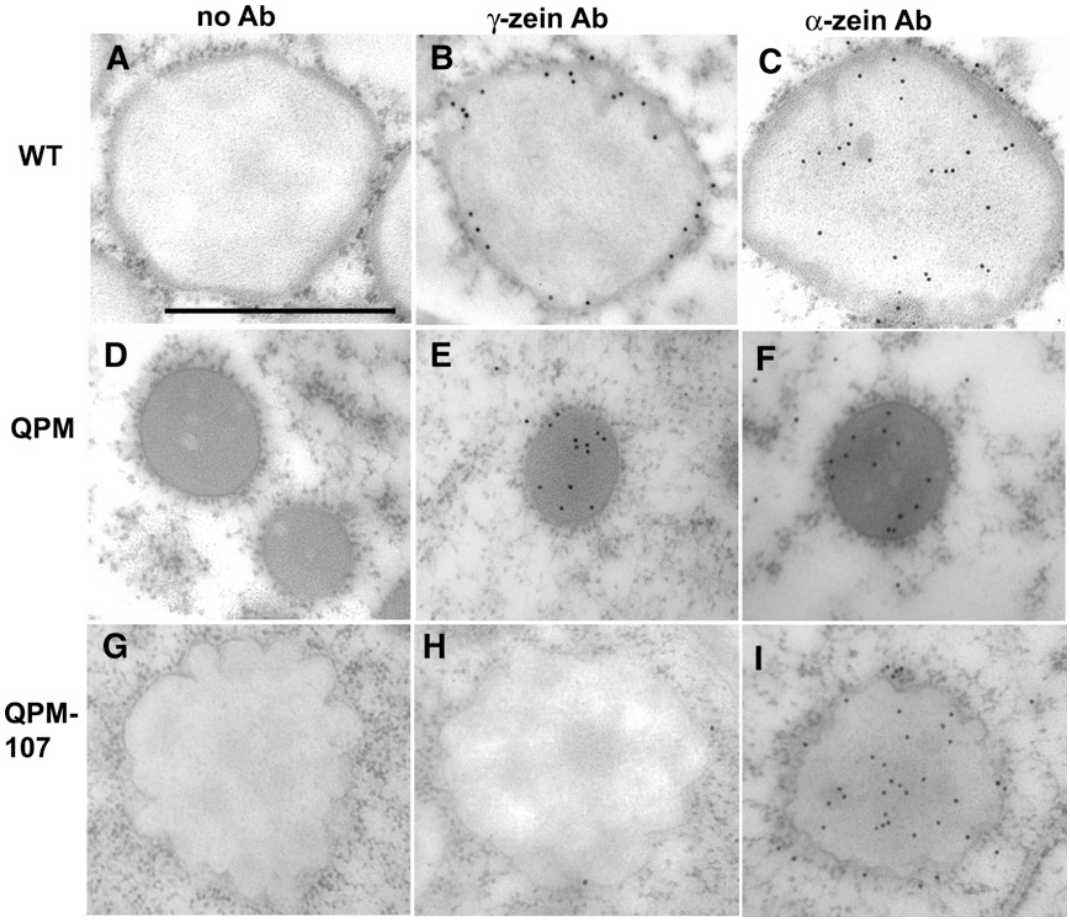


Figure 5. Immunogold TEM analysis showing α- and γ-zein distribution in protein bodies of fourth subaleurone cell layer of 18-DAP endosperm of the wild type, K0326Y QPM, and QPM deletion line 107. Bar = 1 mm. WT, Wild type.

Table I. Protein body number in fourth subaleurone layer of the wild type, K0326Y QPM, and K0326Y line 107

Protein bodies per 100-µm square of fourth subaleurone cell layer were counted from two independent kernels and two independent embedments of each kernel. P values show that counts are significantly different from the wild type.

	W64A Wild Type	K0326Y QPM	K0326Y Line 107
Protein body no.	50.0	87.6	6.2
SD	±3.0	±8.4	±0.8
P value	–	$1.6\text{x}10^{-5}$	$4.4\text{x}10^{-9}$

Table II. Lys and Trp concentrations in mature kernels

Mature Kernel amino acid concentrations for Lys and Trp are shown in grams per $100g$ protein $\pm SD$ (n= 3)

	W64A Wild Type	W64A$o2$	K0326Y QPM	K0326Y Line 107 Qpaque	Ko326Y 107 Semiopaque
Lys	2.39 ± 0.08	4.96 ± 0.04	4.09 ± 0.00	4.93 ± 0.17	4.42 ± 0.01
Trp	0.61 ± 0.05	1.00 ± 0.00	0.92 ± 0.00	0.97 ± 0.13	0.92 ± 0.05

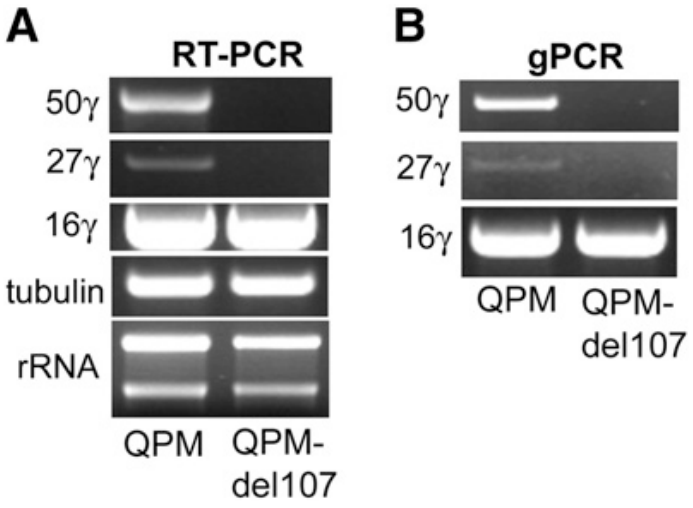


Figure 6. A, RT-PCR showing γ-zein transcript accumulation. B, Genomic PCR analysis of γ-zein genes.

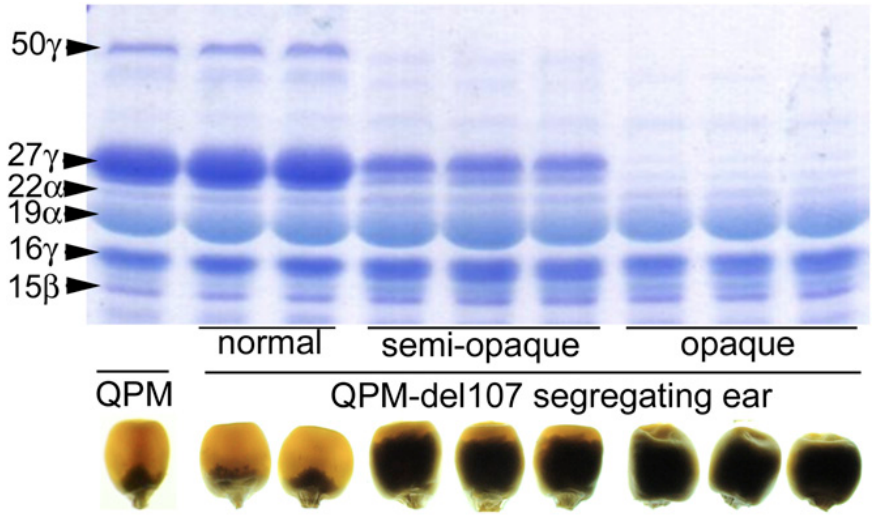


Figure 7. SDS-PAGE and kernel opacity analysis showing haploinsufficiency of γ-zein for endosperm modification. Selected kernels from a segregating ear are shown.

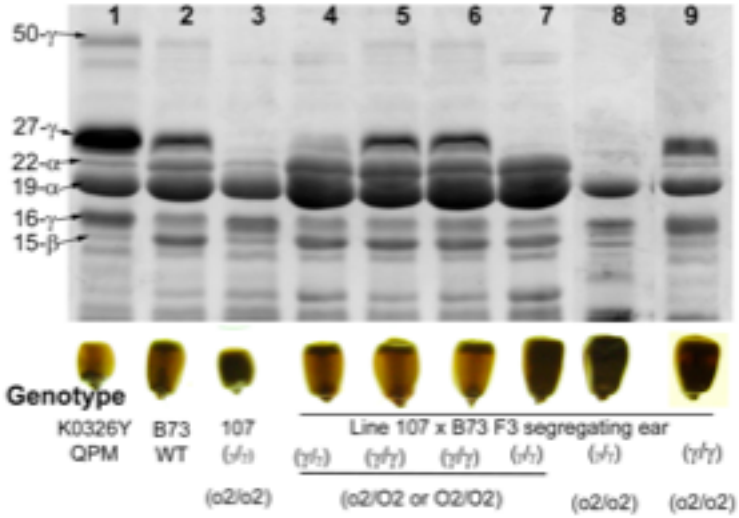


Figure 8. SDS-PAGE showing zein accumulation and its relationship to kernel vitreousness in F3 kernels resulting from a line 107xB73 wild type out-cross.

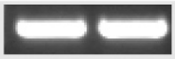
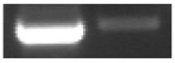
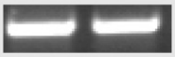
Gene ID	Predicted function	Exon coordinates	Illumina reads	Genomic PCR	RT-PCR
GRMZM2G117230	Alanine aminotransferase	120108726~120108800 120108912~120109115	14:8 14:8		
GRMZM2G138689	50-kDa γ-zein	120173851~120175165	193:2		
GRMZM2G138727	27-kDa γ-zein	120200894~120203011 120201746~120202137 120202221~120203011	86:0 19:0 66:0		
GRMZM2G052667	EREBP-like protein	120351650~120352283 120353107~120354274 120353313~120354274	13:0 19:0 19:0		
GRMZM2G446625	hypothetical protein	120541223~120541408 120543049~120543471 120544243~120546286	25:0 42:0 205:2		
GRMZM2G011463 *	auxin-responsive SAUR family	120711575~120712540	7:0		
GRMZM2G399644 *	small auxin up RNA1	120720817~120721489	17:0		
GRMZM2G099049 *	auxin-repressed protein	120764736~120765217 120764736~120765655 120765550~120765655	1:0 6:0 2:0		
GRMZM2G147966 *	respiratory burst oxidase homolog protein E	121011986~121012145 121010807~121011842 121015111~121015185	23:0 16:0 9:0		
GRMZM2G169201	NADPH oxidase	121019843~121019998 121021786~121022228 121022480~121022991	28:0 50:0 29:0		
GRMZM2G135763	Poly-galacturonase	121252254~121252957 121254902~121255055 121255433~121256304	27:1 26:0 56:2		
GRMZM2G318319	BRCA1 C terminus domain protein	121405051~121405156 121405256~121405518 121405772~121405903	28:0 41:0 27:0		
GRMZM2G429842	subtilisin homologue	121576799~121577612	29:23		

Figure 9. Exon-seq and genomic and RT-PCR verification of deletion on chromosome 7.02. Illumina reads column and PCR columns show non-mutagenized K0326Y versus line 107. Most gene identifications show absence of Illumina reads in more than one exon (denoted by multiple sets of coordinates). GRMZM2G117230 and 429842 are the 59 and 39 flanking intact genes. Genes with an asterisk are either low-abundance transcripts, endosperm non-expressed genes, or pseudogenes.

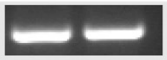
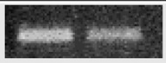
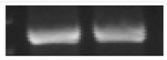
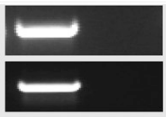
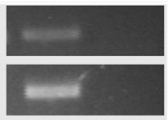
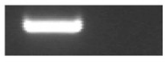
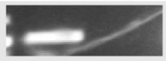
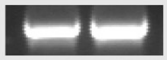
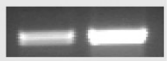
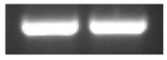
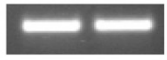
Gene ID	Predicted function	Exon coordinates	Illumina reads	Genomic PCR	RT-PCR
GRMZM2G398628	unknown	49068606~49069246 49069377~49069481 49069516~49069538 49069732~49069782 49069916~49070000 49070243~49070310 49070705~49070954 49071037~49071242	6:6 1:2 0:0 13:2 1:4 22:17 4:1 0:0		
GRMZM2G435380 *	Polygalacturon-ase activity	49132685~49133047 49133156~49133346 49133433~49133640 49133837~49134016 49134122~49134256 49134481~49134663	16:2 9:3 6:3 0:0 2:0 0:0		
GRMZM2G099648	ubiquitin ligase complex	49177017~49178036 49178335~49179500 49185658~49185658 49185904~49186089 49187841~49187932 49188881~49189071	135:1 157:0 10:0 13:0 1:0 0:0		
GRMZM2G040605 *	Clone 323008 mRNA sequence	49308671~49309734	26:0		
GRMZM2G583994	unknown	49376092~49376278	0:0		
GRMZM2G169098	Unknown	49378158~49378477 49385304~49385415 49385522~49385558	0:0 0:0 0:0		
GRMZM2G319760	ATP binding	49408086~49408985 49385558~49396708 49397325~49397325 49397540~49397641 49397931~49398029 49398392~49398526 49398646~49398724 49403154~49403229 49403313~49403402 49406846~49406924 49407025~49407104 49407185~49407259	160:0 5:0 5:0 1:0 1:0 9:0 1:0 22:0 10:0 7:0 13:0 1:0		
Intergenic region	NA	49377028~49377656			
Intergenic region	NA	49385750~49386351			
Intergenic region	NA	49395768~49396313			
GRMZM2G499214	Unknown	49681108~49681654	0:0		
GRMZM2G108302	unknown	49746399~49746452 49746703~49746943 49747014~49747126 49747208~49747321 49747644~49747839 49747930~49747958 49748370~49748465 49752124~49752366	0:0 1:0 0:1 0:0 0:0 0:1 0:5 0:0		

Figure 10. Exon-seq and genomic and RT-PCR verification of deletion on chromosome 9.03. Illumina reads column and PCR columns show non-mutagenized K0326Y versus line 107. Most gene identifications show absence of Illumina reads in more than one exon (denoted by multiple sets of coordinates). GRMZM2G398628 and 435380 are the two non-deleted genes flanking the 59 end of the deletion, and GRMZM2G499214 and 108302 are the two nearest non-deleted genes flanking the 39 end of the deletion. Genes with an asterisk are either low abundance transcripts, endosperm non-expressed genes, or pseudogenes.

REFERENCES

Bolon YT, Haun WJ, Xu WW, Grant D, Stacey MG, Nelson RT, Gerhardt DJ, Jeddeloh JA, Stacey G, Muehlbauer GJ, et al (2011) Phenotypic and genomic analyses of a fast neutron mutant population resource in soybean. Plant Physiol 156: 240–253

Coleman CE, Clore AM, Ranch JP, Higgins R, Lopes MA, Larkins BA (1997) Expression of a mutant α-zein creates the floury2 phenotype in transgenic maize. Proc Natl Acad Sci USA 94: 7094–7097

Dannenhoffer JM, Bostwick DE, Or E, Larkins BA (1995) opaque-15, a maize mutation with properties of a defective opaque-2 modifier. Proc Natl Acad Sci USA 92: 1931–1935

Dumanovi J, Denic M, Ehrenber L, Bergstra K (1969) Radiation-induced heritable variation of quantitative characters in wheat. Hereditas-Genetiskt Arkiv 62: 221–227

Geetha KB, Lending CR, Lopes MA, Wallace JC, Larkins BA (1991) opaque-2 modifiers increase γ-zein synthesis and alter its spatial distribution in maize endosperm. Plant Cell 3: 1207–1219

Geevers HO, Lake JK (1992) Development of modified opaque-2 maize in South Africa. In ET Mertz, ed, Quality Protein Maize. American Society of Cereal Chemists, St. Paul, pp 49–78

Gibbon BC, Larkins BA (2005) Molecular genetic approaches to developing quality protein maize. Trends Genet 21: 227–233

Guo X, Yuan L, Chen H, Sato SJ, Clemente TE, Holding DR (2013) Nonredundant

function of zeins and their correct stoichiometric ratio drive protein body formation in maize endosperm. Plant Physiol 162:1359–1369

Holding DR, Hunter BG, Chung T, Gibbon BC, Ford CF, Bharti AK, Messing J, Hamaker BR, Larkins BA (2008) Genetic analysis of opaque2 modifier loci in quality protein maize. Theor Appl Genet 117: 157–170

Holding DR, Hunter BG, Klingler JP, Wu S, Guo X, Gibbon BC, Wu R, Schulze JM, Jung R, Larkins BA (2011) Characterization of opaque2 modifier QTLs and candidate genes in recombinant inbred lines derived from the K0326Y quality protein maize inbred. Theor Appl Genet 122:783–794

Holding DR, Larkins BA (2006) The development and importance of zein protein bodies in maize endosperm. Maydica 51: 243–254

Holding DR, Meeley RB, Hazebroek J, Selinger D, Gruis F, Jung R, Larkins BA (2010) Identification and characterization of the maize arogenate dehydrogenase gene family. J Exp Bot 61: 3663–3673

Holding DR, Otegui MS, Li BL, Meeley RB, Dam T, Hunter BG, Jung R, Larkins BA (2007) The maize floury1 gene encodes a novel endoplasmic reticulum protein involved in zein protein body formation. Plant Cell 19:2569–2582

Hossain KG, Riera-Lizarazu O, Kalavacharla V, Vales MI, Maan SS, Kianian SF (2004) Radiation hybrid mapping of the species cytoplasmspecific (scsae) gene in wheat. Genetics 168: 415–423

Jia M,Wu H, Clay KL, Jung R, Larkins BA, Gibbon BC (2013) Identification and characterization of lysine-rich proteins and starch biosynthesis genes in the opaque2 mutant by transcriptional and proteomic analysis. BMC Plant Biol 13: 60

Johri MM, Coe EH Jr (1983) Clonal analysis of corn plant development. I. The development of the tassel and the ear shoot. Dev Biol 97: 154–172

Kim CS, Gibbon BC, Gillikin JW, Larkins BA, Boston RS, Jung R (2006) The maize Mucronate mutation is a deletion in the 16-kDa γ-zein gene that induces the unfolded protein response. Plant J 48: 440–451

Kim CS, Hunter BG, Kraft J, Boston RS, Yans S, Jung R, Larkins BA (2004) A defective signal peptide in a 19-kD α-zein protein causes the unfolded protein response and an opaque endosperm phenotype in the maize De*-B30 mutant. Plant Physiol 134: 380–387

Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10: R25

Li X, Song Y, Century K, Straight S, Ronald P, Dong X, Lassner M, Zhang Y (2001) A fast neutron deletion mutagenesis-based reverse genetics system for plants. Plant J 27: 235–242

Lopes MA, Larkins BA (1991) γ-Zein content is related to endosperm modification in Quality Protein Maize. Crop Sci 31: 1655–1662

Lopes MA, Larkins BA (1995) Genetic analysis of opaque2 modifier gene activity in maize endosperm. Theor Appl Genet 91: 274–281

Lopes MA, Takasaki K, Bostwick DE, Helentjaris T, Larkins BA (1995) Identification of two opaque2 modifier loci in quality protein maize. Mol Gen Genet 247: 603–613

Maddaloni M, Donini G, Balconi C, Rizzi E, Gallusci P, Forlani F, Lohmer S, Thompson R, Salamini F, Motto M (1996) The transcriptional activator Opaque-2 controls the expression of a cytosolic form of pyruvate orthophosphate dikinase-1 in maize

endosperms. Mol Gen Genet 250: 647–654

McCarty DR, Settles AM, Suzuki M, Tan BC, Latshaw S, Porch T, Robin K, Baier J, Avigne W, Lai JS, et al (2005) Steady-state transposon mutagenesis in inbred maize. Plant J 44: 52–61

Mertz ET, Bates LS, Nelson OE (1964) Mutant gene that changes protein composition and increases lysine content of maize endosperm. Science 145: 279–280

Myers AM, James MG, Lin QH, Yi G, Stinard PS, Hennen-Bierwagen TA, Becraft PW (2011) Maize opaque5 encodes monogalactosyldiacylglycerol synthase and specifically affects galactolipids necessary for amyloplast and chloroplast function. Plant Cell 23: 2331–2347

Naito K, Kusaba M, Shikazono N, Takano T, Tanaka A, Tanisaka T, Nishimura M (2005) Transmissible and nontransmissible mutations induced by irradiating Arabidopsis thaliana pollen with γ-rays and carbon ions. Genetics 169: 881–889

Poethig RS, Coe EH, Johri MM (1986) Cell lineage patterns in maize embryogenesis: a clonal analysis. Dev Biol 117: 392–404

Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26: 139–140

Robinson MD, Oshlack A (2010) A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol 11: R25

Sato Y, Shirasawa K, Takahashi Y, Nishimura M, Nishio T (2006) Mutant selection from progeny of γ-ray-irradiated rice by DNA heteroduplex cleavage using Brassica petiole extract. Breed Sci 56: 179–183

Schmidt RJ, Burr FA, Aukerman MJ, Burr B (1990) Maize regulatory gene opaque-2 encodes a protein with a “leucine-zipper” motif that binds to zein DNA. Proc Natl Acad Sci USA 87: 46–50

Settles AM (2005) Maize community resources for forward and reverse genetics. Maydica 50: 405–414

Settles AM, Holding DR, Tan BC, Latshaw SP, Liu J, Suzuki M, Li L, O’Brien BA, Fajardo DS, Wroclawska E, et al (2007) Sequence-indexed mutations in maize using the UniformMu transposon-tagging population.BMC Genomics 8: 116

Vasal SK, Villegas E, Bjarnason M, Gela W, Goertz P (1980) Genetic modifiers and breeding strategies in developing hard endosperm *opaque2* materials. In WG Pollmer, RH Phillips, eds, Quality Traits of Maize for Grain and Silage Use. Martinus Nijhoff, London, pp 37–73

Vollbrecht E, Duvick J, Schares JP, Ahern KR, Deewatthanawong P, Xu L, Conrad LJ, Kikuchi K, Kubinec TA, Hall BD, et al (2010) Genome-wide distribution of transposed Dissociation elements in maize. Plant Cell 22:1667–1685

Wallace JC, Lopes MA, Paiva E, Larkins BA (1990) New methods for extraction and quantitation of zeins reveal a high content of γ-zein in modified opaque-2 maize. Plant Physiol 92: 191–196

Wang GF, Wang F, Wang G, Wang F, Zhang XW, Zhong MY, Zhang J, Lin DB, Tang YP, Xu ZK, et al (2012) Opaque1 encodes a myosin XI motor protein that is required for endoplasmic reticulum motility and protein body formation in maize endosperm. Plant Cell 24: 3447–3462

Weil C, Monde R, Till B, Comai L, Henikoff S (2005) Mutagenesis and functional

genomics in maize. Maydica 50: 415–424

Wu JL, Wu CJ, Lei CL, Baraoidan M, Bordeos A, Madamba MRS, Ramos-Pamplona M, Mauleon R, Portugal A, Ulat VJ, et al (2005) Chemicaland irradiation-induced mutants of indica rice IR64 for forward and reverse genetics. Plant Mol Biol 59: 85–97

Wu Y, Holding DR, Messing J (2010) γ-Zeins are essential for endosperm modification in quality protein maize. Proc Natl Acad Sci USA 107:12810–12815

Wu Y, Messing J (2010) RNA interference-mediated change in protein body morphology and seed opacity through loss of different zein proteins. Plant Physiol 153: 337–347

Wu Y, Yuan L, Guo X, Holding DR, Messing J (2013) Mutation in the seed storage protein kafirin creates a high-value food trait in sorghum. Nat Commun 4: 2217

Yuan L, Dou Y, Kianian SF, Zhang C, Holding DR (2014) Deletion mutagenesis identifies a haploinsufficient role for gamma-zein in opaque2 endosperm modification. Plant Physiol 164: 119-130

CHAPTER 3

CHARACTERIZATION OF K0326Y DELETION LINE 198: A QPM OPAQUE REVERTANT WITH IMPROVED PROTEIN QUALITY

ABSTRACT

A K0326Y-QPM deletion mutant, called line 198, has decreased abundance of all zeins, especially the 19-kD α-zeins. Like line 107, the soft starchy endosperm of line 198 could not transmit light and therefore the kernels have an opaque phenotype. We measured lysine and tryptophan concentration in mature kernels of the line 198 and compared it with the K0326Y QPM progenitor, the W64A wild type, and the W64A*o2*. Lysine concentration in 198 kernels was 12% higher than that of *W64Ao2,* which suggests that it could be a useful breeding resource for the production of high-lysine maize varieties. Genomic PCR, RT-PCR and Southern blot analysis did not indicate a deletion of one or more Z1 gene subfamilies, and suggested that the decreased γ- and β-zeins abundance results from a secondary effect. Bulked Segregant Analysis was used to map the mutation to a ~30 Mb region on chromosome 3.02 which does not contain α-zein gene families, which further ruled out a physical zein gene deletion in the genome. Putative small deletions identified from Illumina exon-seq data were determined to be false positives by genomic PCR and RT-PCR. Also exon- and RNA-seq data did not show any single or multiple gene deletion in the ~30 Mb mapping interval, which suggested that the mutation is possibly a small deletion within a gene or in an intergenic regulatory region that is not identifiable with exon- or RNA-seq. We are conducting whole genome sequencing using HiSeq 2500 to pinpoint the mutation of line 198. The characterization of line 198 is on-going and may identify a novel regulatory factor that directly or indirectly affects the amount of 19-kD α-zeins and is allowing us to explore its specific role in normal vitreous endosperm formation.

METHODS

The methods for radiation dosage testing, γ-radiation mutagenesis, screening of M2, heritability testing in M3, genotyping, RT-PCR, zein protein analysis, exon- and RNA-seq are described in Chapter 2 (Yuan et al., 2014).

Bulked Segregant Analysis (BSA) mapping

BSA mapping was conducted at the Genomic Technologies Facility of Iowa State University (ISU). The mapping population originated from crossing the homozygous line 198 with B73 wild type. After self-pollinating the F1 and F2 plants, the segregating F3s were classified into opaque and vitreous groups based on their kernel phenotype. The number of individuals in each bulk was 30 mutants and 17 wild types. Leaf tissues from all line 198 (or wild type) individuals were pooled and a single genomic DNA isolation was performed using a modified urea extraction protocol (Holding et al., 2008). DNA concentration and purity was assayed with Thermo NANODROP 1000 and 100μg crude DNA was purified using the Qiagen DNeasy Plant Mini Kit. 10μg samples of purified DNA were used for BSA. At the ISU Genomic Technologies Facility, the two purified DNA pools were subjected to a series of multiplexed Sequenom assays. 1016 of the co-dominant and dominant markers were previously successfully mapped to B73 × Mo17 (IBM) recombinant inbred line (RIL) and 500 validated SNP markers were finally used for our two DNA pools analyses. The possible locations of the genes responsible for the mutant phenotype were identified by the analysis of the resulting SNP-typing data. In our experiment, linkage was considered significant if the logarithm 2 score was more than 2 (Liu et al., 2010).

RESULTS

Identification of line 198 with general zein decrease and further improved protein quality

Of the 293 QPM deletion families, line 198 is one of the opaque mutants that we are characterizing. Like line 107, line 198 exhibited mutant kernels in the M2 ears, which indicates dominance or semi-dominance of the mutation, because of the ontogeny of the ear and tassel initial cells in the embryonic shoot apical meristem (Johri and Coe Jr, 1983; Poethig et al., 1986). Segregating and homozygous M3 ears of this line are shown in Figure 1. Starchy mutant kernels cannot transmit light and therefore, show an opaque phenotype. As described in Chapter 2, although the phenotypes of M2 ears was potentially affected by environmental factors, the presence of opaque kernels in well filled M2 ears of line 198 suggested that line 198 is a dominant mutant. This was verified by its phenotypic heritability in the M3 generation. Analysis of zein proteins of line 198 revealed a general decreased abundance in all zein fractions, especially the 19-kD α-zeins (Figure. 2A), which was confirmed with immunoblot analysis (Figure. 2B). Notably, in line 198 the 19-kD α-zein is detectable only after longer exposure. 19-kD α-zeins were significantly reduced in *o2* and QPM compared with the W64A wild type but not to the same extent as the 22-kD α-zeins. However, the 19-kD α-zeins in K0326Y QPM were dramatically further less abundant in K0326Y line 198 (Figure. 2A and 2B). The correlation between the amount of 19-kD α-zeins and the loss of kernel vitreousness in 198, and the potential to identify a physical α-zein deletion or a novel, non-O2, α-zein regulatory factor prompted us to characterize this mutant further.

We measured the lysine and tryptophan concentration in mature kernels of the line 198 and compared it with the K0326Y QPM progenitor, W64A wild type and W64A*o2* (Table 1). As we described in the characterization of line 107, comparison of W64A WT and *o2* kernels showed the expected levels of *o2* increase for these amino acids. Increase in the lysine and tryptophan percentage in QPM was not as pronounced as W64A*o2* probably as a result of higher-level accumulation of lysine-devoid 27-kD γ-zein. Notably, the lysine concentration in line 198 was even 12% higher than that in W64A*o2*.

Line 198 has no deletions in sub-families of α-zeins genes or in γ/β –zein genes

Given that line 198 mutant kernels have dramatically reduced zein abundance, we sought to determine if the protein decrease resulted from deletion of a zein regulatory factor, a physical genomic deletion of zein gene(s) or an indirect effect on zein synthesis. As described in Chapter 1, α-zein genes include four gene sub-families: *Z1A, Z1B, Z1C* and *Z1D*. *Z1A*, *Z1B* and *Z1D* sub-families encode the 19-kD α-zeins and *Z1C* sub-family encodes 22-kD α-zeins. The high sequence similarities of α-zein gene members in each sub-family mean that PCR primers that specifically amplify individual family members cannot be designed. However, it is possible to design primers that can distinguish between expression of α-zein subfamilies with PCR (Feng et al., 2009). Therefore, we used family specific primers for each α-zein gene sub-family to screen for deletions of whole gene sub-families. By genomic PCR, we identified no such α-zein deletions (Figure. 3A). Conventional semi-quantitative RT-PCR showed that the transcript levels of α-zeins genes, especially Z1B, Z1C and Z1D groups, were lower in line 198 compared to control (Figure. 3B). Quantitative real time RT-PCR in developing kernels of line 198

showed drastic decrease in expression of all α-zein genes (Figure. 4, A-D) and general slight reduction in expression of γ-zein genes (Figure. 4, E-H). Z1A sub-family transcripts were approximately 4% of the level of the QPM control (Figure. 4A). The transcripts of Z1B, Z1C and Z1D sub-families were less than 1% of the level of the QPM control (Figure. 4, B-D). The transcripts of 50- 27-, 16-kD γ-zein and 15-kD β-zein were reduced to approximately 80%, 70%, 45% and 4% of the control level respectively (Figure. 4, E-H). Given that the 50-, 27- and 16-kD γ-zein and the 15-kD β-zein are encoded by single genes, we propose that there are no genomic deletions in γ- and β-zein genes and their decrease results from an indirect effect.

Southern blot analysis did not suggest a deletion encompassing Z1 gene subfamilies of line 198

PCR analysis did not suggest deletions of whole sub-families of α-zeins in line 198. However, that data cannot rule out the possibility that only one gene or a partial region of a α-zein sub-family region was deleted. The B73 reference line has 41 genes that encode α-zeins and they are located in six chromosomal regions. To further address the possibility of physical genomic deletions in zein gene regions that resulted in α-zeins decrease, we conducted Southern blot analysis using a mixed probe amplified from all α-zein sub-families. No missing band was found in DNA from line 198 leaves, which is also not supportive of a large physical α-zein gene deletion.

Bulked Segregant Analysis (BSA) indicated a likely map position for the line 198 mutation on chromosome 3.02

BSA as an efficient mapping strategy is conducted by screening for marker differences between two-bulked DNA pools derived from a segregating population that originated from a single cross between a mutant line and another inbred line with a different genetic background. We utilized the Genomic Technologies Facility at Iowa State University to map line 198. To map the mutation causing the kernel phenotype, line 198 was crossed with B73. After self-pollinating the F1 and F2 plants, the segregating F3s were classified into opaque and vitreous groups based on their kernel phenotype. The number of individuals in each bulk was 30 mutants and 17 wild types. Tests on the effects of pool size on the ability to identify a mutation showed that ~20 individuals in the mutant pool was sufficient (Liu et al., 2010) (Holding et al., 2008). Our BSA mapping result showed two strong signal peaks on chromosome 3 and 7 respectively (Figure. 6). 11 SNP markers on chromosome 3 were considered significant and expanded a genomic region from coordinates 168,231,982 to 199,026,446. Fifteen significant SNP markers on chromosome 7 expanded a genomic region from coordinates 4,723,834 to 154,268,338. Since mutant line 198 was created from QPM, which contains the *o2* mutation, and unmodified *o2* mutant kernels may also be segregating in F2 and F3 ears, the peak on chromosome 7 almost certainly corresponds to *opaque-2*. We cannot rule out the possibility that line 198 may also map within this chromosome 7 region. However, the peak on chromosome 3 likely corresponds to the mutation in line 198. Based on the marker data producing the peak on chromosome 3, we mapped the mutation to a ~30 Mb region of chromosome 3.02 where no α-zeins families reside. In addition, the filtered exon-seq data of line 198, across the genome, a total 41 candidate genes were identified and tested by genomic and RT-PCR but were determined to be false positive. Those

exons were called deletion candidates since they meet the criterion we set that P-value was smaller than 10^{-5}. And the number of reads for the mutagenized line were not larger than 0, 1, 2, 3, and 5 when the numbers of reads for the non-mutagenized reference line were in ranges of (30-50), (50-70), (70-100), (100-150), (150 and above), respectively. The BSA result helped us narrow down the potential deletion region from the broad whole genome. Further, based on the BSA data, we identified all the candidate genes located in the narrow chromosome 3.02 region from exon-seq data. This region, containing 2052 genes in total, does not show any large deletions (data not shown). This may suggest the mutation in line 198 is a small deletion within a gene or a deletion in an intergenic region in chromosome 3.02 that contains regulatory elements.

DISCUSSION

α-zeins are essential zein fractions in maize endosperm. They account for more than 70% of total zein protein. The two subfamilies, 22-kD α-zeins and 19-kD α-zeins, comprise more than 20% and 40% of the total zein respectively (Huang et al., 2006), which indicates the 19-kD α-zeins are a major component. 27-kD γ-zein is one of the essential modifiers for vitreous endosperm formation and 50-kD γ-zein was also proposed to be involved as an *o2* modifier (Wu et al., 2010; Yuan et al., 2014). However, the role of 19-kD α-zeins in endosperm modification in QPM and their essentiality in normal vitreous endosperm formation is not known. Opaque line 198 has dramatically reduced level of 19-kD α-zein accumulation compared with non-mutagenized QPM. In addition, the lysine concentration in line 198 was 12% higher than that in W64A*o2*, which suggested this mutant has higher protein quality. By genomic PCR, RT-PCR and Southern blot

analysis we ruled out the presence of a deletion that encompassed one or more Z1 gene subfamilies. But this led us to further consider the possibility of a deletion on single dominantly expressed member(s) of Z1 gene subfamilies. Even though 19-kD α-zeins comprise 25 members in B73, some of them are not expressed, therefore a deletion in a locus with very high expression could dramatically reduce overall 19-kD α-zein abundance. Our BSA data were not consistent with this possibility, since BSA mapped the mutation to chromosome 3.02 where no α-zein families reside (Figure. 7) (Miclaus et al., 2011). However, the gene cluster of Z1B subfamily of 19-kD α-zeins locates on the identified BSA chromosome 7 region, so it is not completely impossible the strong signal peak on chromosome results from both *O2* and deletion encompassing Z1B subfamily genes. We are conducting full genome resequencing to try to identify the causative deletion in line 198. As described in Chapter 1, O2 is a transcription factor that regulates genes including 22-kD α-zeins, while 19-kD α-zein genes are not regulated by O2. It is possible that an unidentified factor that regulates 19-kD α-zeins is deleted in line 198. It is also possible that the mutation abolishes the vitreous endosperm formation through a mutation that indirectly reduces α-zein abundance but is not directly related to zein genes or their expression. Such indirect effects on zein abundance have been observed in opaque mutants like *mto140*, *o5* and *o7* (Holding et al., 2010; Miclaus et al., 2011; Myers et al., 2011). Even though genome-wide candidate deletions identified using Illumina exon- and RNA-seq were determined to be false positive by genomic PCR and RT-PCR (data not shown), the original exon- and RNA-seq data did not show any large deletion in the ~ 30 Mb region, which suggested the mutation is possibly a small deletion within a gene or gene promoter region in chromosome 3.02.

The characterization of the line 107 (chapter 2) and line 198 demonstrates the feasibility of radiation mutagenesis to study gene and genome function with modern sequencing technologies. As with other mutagenesis methods that create multiple mutations, reverse genetics methods are necessary for proving association between the candidate mutations, which is especially important if a mutant have more than one deletion, or a deletion contains several genes. To identify additional mutant alleles, two different transposon insertion collections can be screened by blast searching. These are the AC/DS collection and the UniformMu collection. If we are successful in identifying the probable causative mutation in line 198, we will try to find a mutant from the above collections. Such a mutant should display the same phenotype as mutant line 198, which will suggest it is this mutation that causes the line 198 phenotype. Allelism tests between an ordered mutant allele and line 198 would confirm a candidate gene if the mutants do not complement each other. In the absence of available additional mutant alleles, RNAi or artificial micro RNA lines can be made. A transgenic plant should display the same phenotype as the line 198. For a gain of function strategy, an over-expression construct can be used to test complementation of the198 mutant phenotype. The characterization of line 198 is ongoing and provides us an opportunity to explore the specific role of 19-kD α-zeins in QPM and normal vitreous endosperm formation and possibly identify a novel regulatory factor for 19-kD α-zeins.

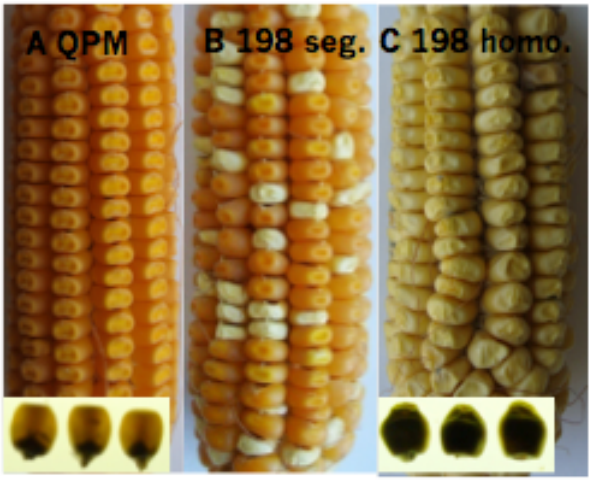


Figure 1. M3 ear phenotypes of K0326Y QPM control ear and segregating and homozygous K0326Y deletion line 198. Inserts in A and C show light box phenotypes.

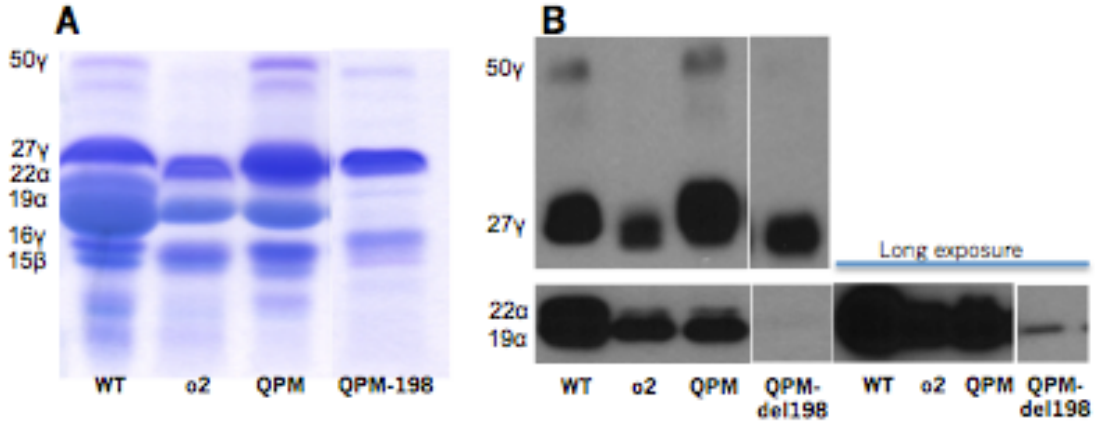


Figure 2. A, SDS-PAGE analysis of zein proteins from mature kernel of deletion line 198 compared with K0326Y QPM and the W64A wild type and *W64Ao2*. B, Western analysis of zein proteins loaded at 1:1,000 dilution of gel shown in A, probed with total α-zein antiserum (1:10,000) and 27-kD γ-zein antiserum (1:2,000).

Table 1. Lys and Trp concentrations in mature kernels of 198

Mature Kernel amino acid concentrations for Lys and Trp are shown in grams per 100g protein ±SD (n = 3)

	W64A Wild Type	*W64AO2*	K0326Y QPM	QPM-de1198
Lys	2.39 ± 0.08	4.96 ± 0.04	4.09 ± 0.00	5.56 ± 0.03
Trp	0.61 ± 0.05	1.00 ± 0.00	0.92 ± 0.00	0.94 ± 0.04

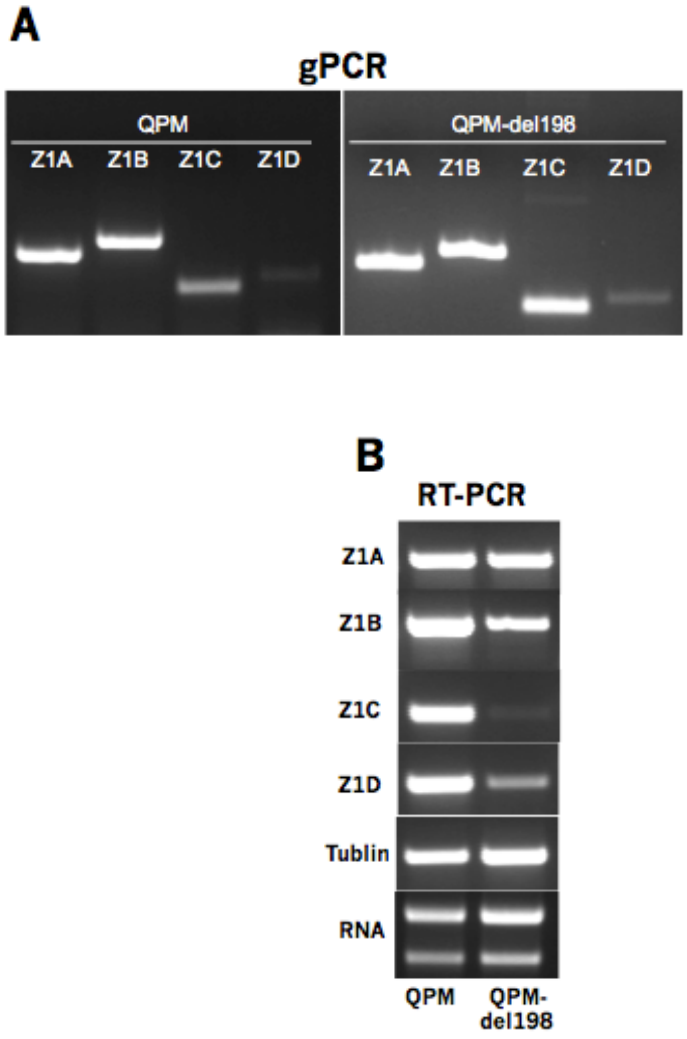


Figure 3. A, Genomic PCR of Z1 locus in both QPM non-mutagenesis line and mutant line 198. B, RT-PCR showing Z1 gene transcript accumulation in 198 compared with QPM non-mutagenesis. Primers are sub-family specific (Feng et al., 2009)and work for both qPCR and RT-PCR.

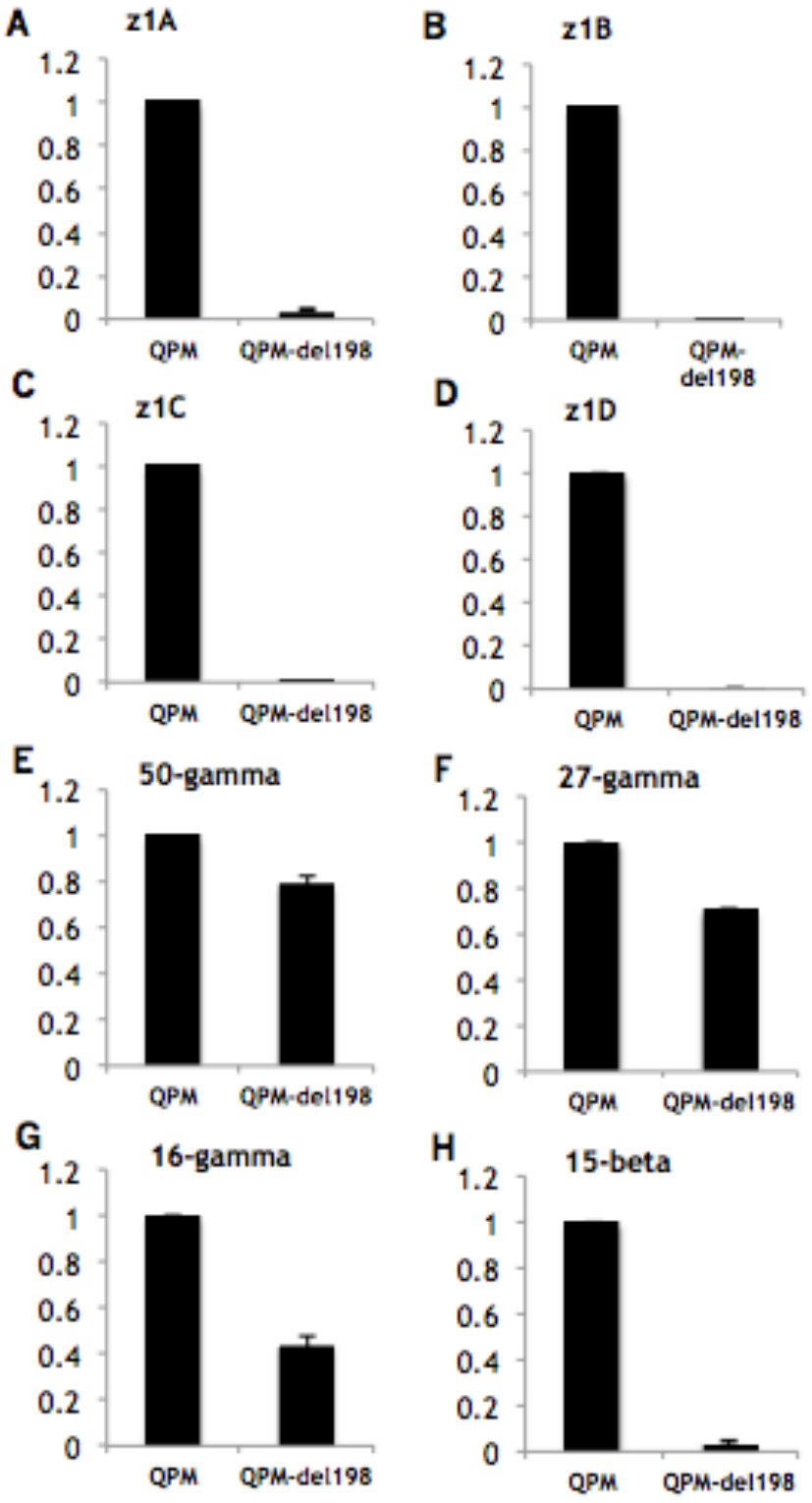


Figure 4. Expression analysis of zein genes by quantitative real time RT-PCR in mutant line 198 relative to QPM non-mutagenesis. Each graph represents the measurement of a different zein gene in this mutant. α-zein primers are sub-family specific whereas others are gene specific.

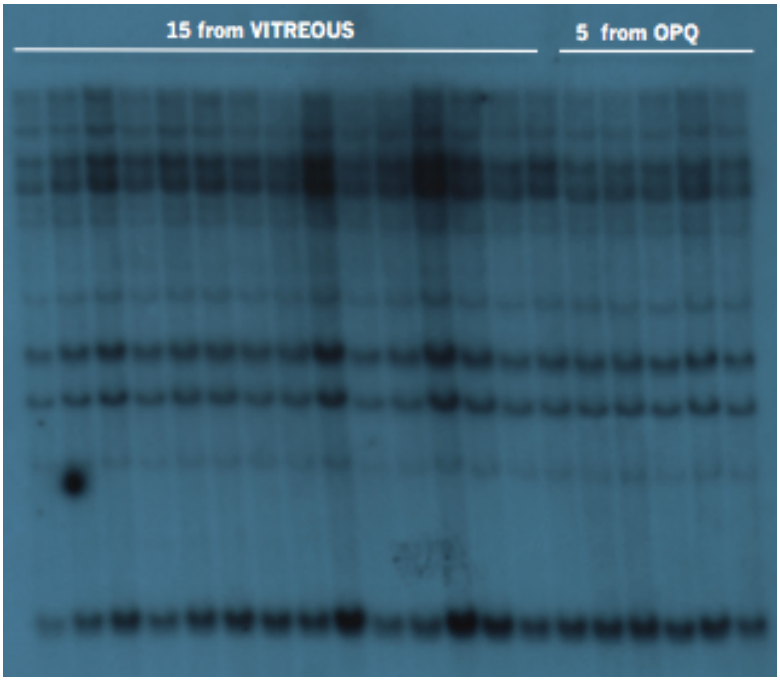


Figure 5. Southern Blot analysis of α-zeins genes in 5 line 198 mutants compared with 15 controls. Probe: mix of z1A B C D zein PCR products amplified from genomic DNA using gene family-specific primers.

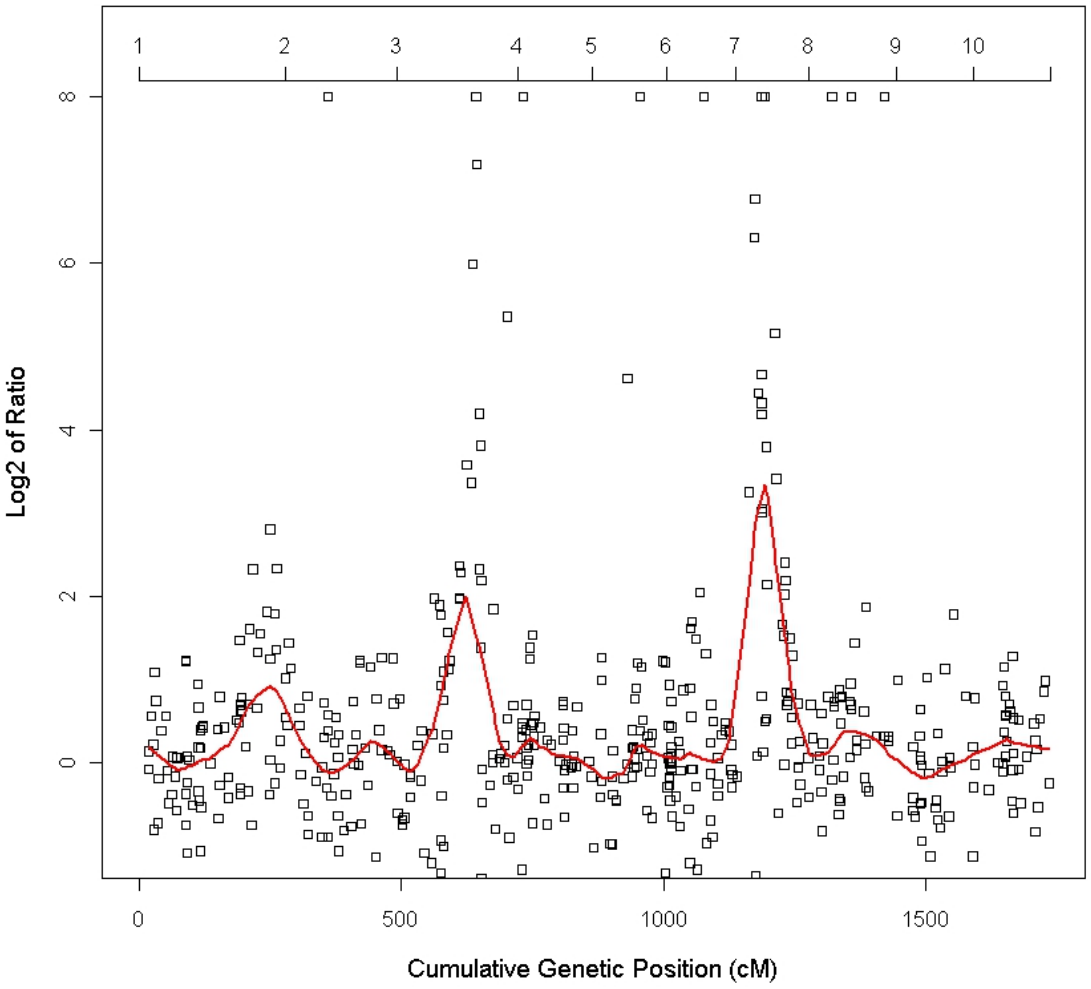


Figure 6. Bulked Segregant analysis mapping of mutation in 198.

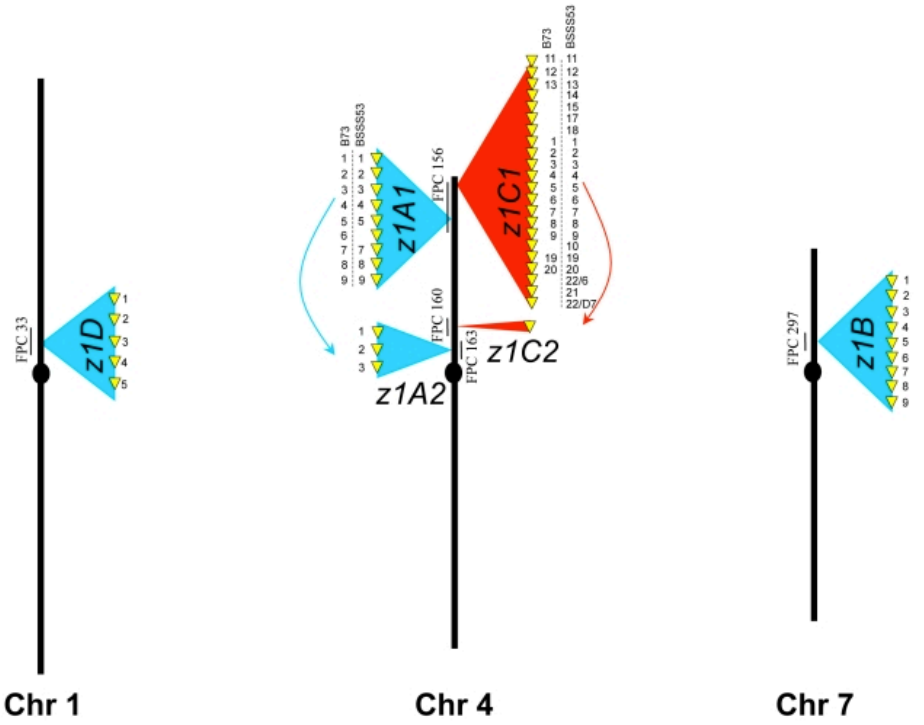


Figure 7. Genomic distribution of α-zeins genes in B73 and BSSS53 inbred lines. 19-kD α-zeins clusters are represented as blue and 22-kD α-zeins clusters are represented as red. Yellow arrows represent zein gene copies at each locus. (Source: Miclaus et al., 2011)

REFERENCES

Feng L, Zhu J, Wang G, Tang Y, Chen H, Jin W, Wang F, Mei B, Xu Z, Song R (2009) Expressional profiling study revealed unique expressional patterns and dramatic expressional divergence of maize α-zein super gene family. Plant molecular biology 69: 649-659

Holding DR, Hunter BG, Chung T, Gibbon BC, Ford CF, Bharti AK, Messing J, Hamaker BR, Larkins BA (2008) Genetic analysis of opaque2 modifier loci in quality protein maize. Theoretical and Applied Genetics 117: 157-170

Holding DR, Meeley RB, Hazebroek J, Selinger D, Gruis F, Jung R, Larkins BA (2010) Identification and characterization of the maize arogenate dehydrogenase gene family. J Exp Bot 61: 3663-3673

Huang S, Frizzi A, Florida CA, Kruger DE, Luethy MH (2006) High lysine and high tryptophan transgenic maize resulting from the reduction of both 19- and 22-kD alpha-zeins. Plant Mol Biol 61: 525-535

Johri M, Coe Jr E (1983) Clonal analysis of corn plant development: I. The development of the tassel and the ear shoot. Developmental biology 97: 154-172

Liu S, Chen HD, Makarevitch I, Shirmer R, Emrich SJ, Dietrich CR, Barbazuk WB, Springer NM, Schnable PS (2010) High-throughput genetic mapping of mutants via quantitative single nucleotide polymorphism typing. Genetics 184: 19-26

Miclaus M, Wu Y, Xu J-H, Dooner HK, Messing J (2011) The maize high-lysine mutant opaque7 is defective in an acyl-CoA synthetase-like protein. Genetics 189: 1271-1280

Miclaus M, Xu J-H, Messing J (2011) Differential gene expression and epiregulation of alpha zein gene copies in maize haplotypes. PLoS genetics 7: e1002131

Myers AM, James MG, Lin Q, Yi G, Stinard PS, Hennen-Bierwagen TA, Becraft PW (2011) Maize opaque5 encodes monogalactosyldiacylglycerol synthase and specifically affects galactolipids necessary for amyloplast and chloroplast function. The Plant Cell Online 23: 2331-2347

Poethig R, Coe Jr E, Johri M (1986) Cell lineage patterns in maize embryogenesis: a clonal analysis. Developmental Biology 117: 392-404

Wu Y, Holding DR, Messing J (2010) γ-Zeins are essential for endosperm modification in quality protein maize. Proceedings of the National Academy of Sciences 107: 12810-12815

Yuan L, Dou Y, Kianian SF, Zhang C, Holding DR (2014) Deletion mutagenesis identifies a haploinsufficient role for gamma-zein in opaque2 endosperm modification. Plant Physiol 164: 119-130

APPENDIX 1

Use of RNA interference for studying prolamin function

As described in chapter 1, RNAi has been used to selectively eliminate zein transcripts, which has advanced our understanding of zein function and the inverse relationship between zein amount and kernel protein quality. But questions concerning their distinct roles remain unanswered. We addressed the extent of functional non-redundancy within the γ- and α-zein subclasses using RNAi technology (Guo et al., 2013). My involvement in this work was the in depth analysis of transcript and protein levels in endosperms of these lines.

In our lab, we used three RNAi transgenes to target all α-zeins, a 27-kD γ-zein, and other γ-zein classes (16-kD γ-, 50-kD γ-, and 15-kD β-zeins) (Figure. 1). We also crossed these RNAi events to derive the pairwise and triplicate combinations to test combinations that resulted in kernel opacity and to test the additive effects of zein depletion at the transcript and protein levels.

Our strategy for complete knockdown of 19- and 22-kD α-zeins involved a synthetic gene composed of approximately 220-bp conserved regions of a highly expressed member of each of the 19-kD α-zein gene subfamilies (*Z1A*, *Z1B*, and *Z1D*) and the 22-kD subfamily (*Z1C*). Endosperm specific expression of this RNAi cassette was driven by the 27-kD γ-zein promoter (Figure. 1A). To measure the effect of the RNAi transgene on endogenous α-zein genes, we used quantitative real-time reverse transcription (qRT)-PCR of

endosperm RNA from kernels that were genotyped by PCR using DNA extracted from their dissected embryos. The α-zein RNAi line had transcript abundance that was reduced to less than 1% of the level of 19-kD subclass transcripts (*Z1A* and *Z1B*) in the wild-type control (Figure. 2A and 2B) and *Z1D* at approximately 2% to 5% of the wild-type control (Figure. 2D). The 22-kD subclass transcripts (*Z1C*) were reduced to approximately 10% of the wild-type level (Figure. 2C). These reductions were observed in both the single α-zein RNAi line and in combination with other RNAi constructs. Slight non-specific reduction in expression of *Z1A*, *Z1B*, *Z1C*, and *Z1D* class transcripts also resulted from expression of both γ-zein RNAi constructs described below (Figure. 2A-D). This may indicate that high-level expression of the γ-zein RNAi constructs in the endosperm can indirectly reduce endogenous expression of α-zein through competition for transcriptional machinery. Alternatively, improper protein body formation caused by γ-zein RNAi-mediated defects may cause negative feedback on expression of α-zein genes.

γ-zein family members 27-kD, 50-kD, and 16-kD γ-zein and the 15-kD β-zein share a common location in the protein body periphery and are assumed to have redundant function. To address the functional redundancy of 27-kD, 50-kD, and 16-kD γ-zein and the 15-kD β-zein, we made transgenic lines containing two different RNAi constructs. The first of these used a synthetic gene composed of 212-bp segments of the 50-kD and 16-kD γ-zein gene and the 15-kD β-zein gene (Figure. 1B). This construct, whether alone or in combination with other RNAi constructs, reduced the abundance of all of the three targeted γ-zein proteins to very low levels, usually less than 1% of the wild-type level (Figure. 2F-H). This construct also reduced the level of the 27-kD γ-zein to

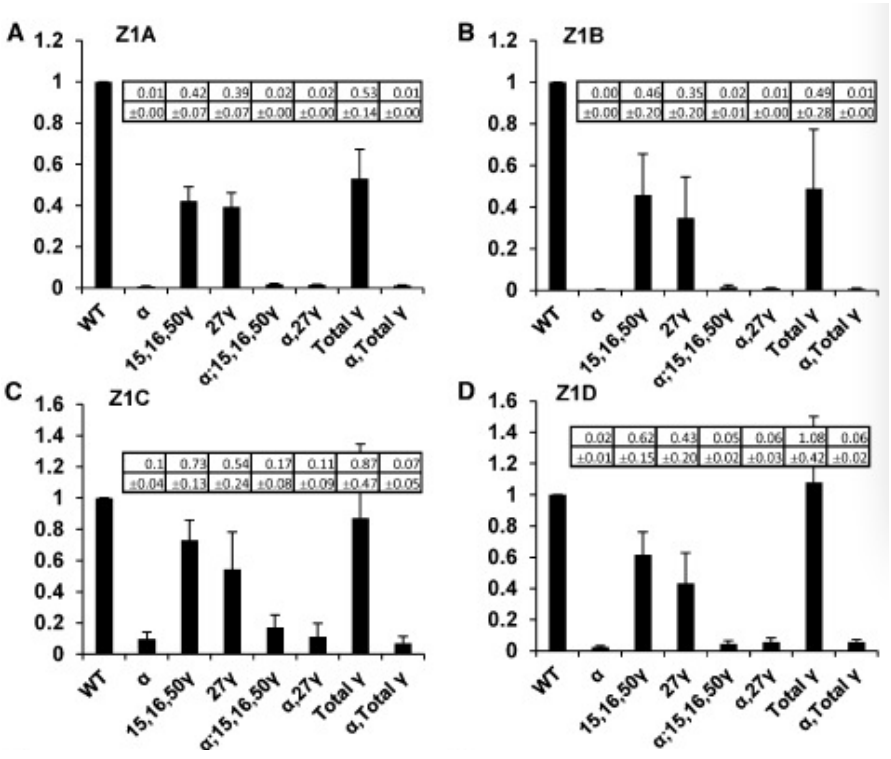
approximately 20% to 30% of the wild-type level (Figure. 2E), probably due to the similarity of the 16-kD and 27-kD γ-zein. Similar cross repression was previously described (Wu and Messing, 2010).

We specifically targeted 27-kD γ-zein for knockdown using an RNAi construct that would form a hairpin loop that was composed of the entire 27-kD γ-zein open reading frame (Figure. 1C). Expression of this transgene resulted in 1% or less of the wild-type transcript level for both 27-kD and 16-kD γ-zein (Figure. 2E and F). This is probably because of the sequence similarity between the 27-kD and 16-kD γ-zein. And expression of this transgene also partially suppressed the 50-kD γ-zein to 46% of the wild type (Figure. 2G) and the 15-kD β-zein to 12% of the wild type (Figure. 2H). The basis for the transcript reductions of 50-kD γ-zein and 15-kD β-zein is unknown. However, when the 27-kD γ-zein RNAi was combined with the α-zein RNAi, while the 27-kD γ-zein remains suppressed, the 16-kD γ-zein suppression is largely lifted. A similar effect was observed with the 50-kD γ-zein, whose transcript was increased in both of the γ-zein RNAi construct lines when combined with the α-zein RNAi. Together, these suggest that with reduced α-zein synthesis, non-specific effects of other RNAi events may be less pronounced because of increased availability of translational resources.

To investigate protein body formation, transgenic kernels at 18 DAP were used for low-magnification TEM for analysis of protein body density and high-magnification TEM for analysis of protein body size and morphology (Figure. 3 and 4). TEM results showed that the α-zein RNAi event resulted in protein bodies that were around 25%-30% of normal

size, but there was no change in protein body shape or number, which suggests that α-zeins are not necessary for protein body initiation but contribute to protein body expansion (Figure. 3B and 4B). The 27-kD γ-zein RNAi did not reduce protein body size, but induced an undulated protein body shape (Figure. 3C and 4C), which suggests that the 27-kD γ-zein is not necessary for protein body filling, but plays a role in packaging α-zein. However, the 27-kD γ-zein event caused a significantly reduced protein body number (Figure. 3C), which suggests the 27-kD γ-zein plays an essential role in protein body initiation. The kernels with both 27-kD γ-zein and α-zein RNAi have similar severe loss of protein body number to that of 27-kD γ-zein RNAi alone (Figure. 3C and F), which suggests the 27-kD γ-zein plays a major role in protein body initiation. The 16/50/15-kD γ-zeins RNAi construct caused the protein body size to be reduced by 50%, but protein bodies had normal shape and number (Figure. 3D and 4D), which suggests that some or all 16/50/15-kD γ-zeins are necessary for protein body expansion and they are not involved in protein body initiation. By combining all three RNAi constructs, we achieved low levels of all zeins. Protein body number was the lowest in this event, but protein body size and morphology ranged from slightly reduced to normal (Figure. 3H and 4H). This is somewhat surprising and suggests that protein bodies of normal size and shape can form so long as the different zeins are in approximately correct stoichiometric ratio. RNAi effects on protein body number, size, and morphology suggest that there is significant non-redundancy of the roles of zeins in protein body formation.

Figure 1. Diagram of zein RNAi constructs with 27-kD γ-zein promoter (p27γ) and Cauliflower mosaic virus terminator (T35S). (Source: Guo et al., 2013) A, α-zein RNAi construct. B, 50-kD γ-, 16-kD γ-, and 15-kD β-zein RNAi construct. C, 27-kD γ-zein RNAi construct.



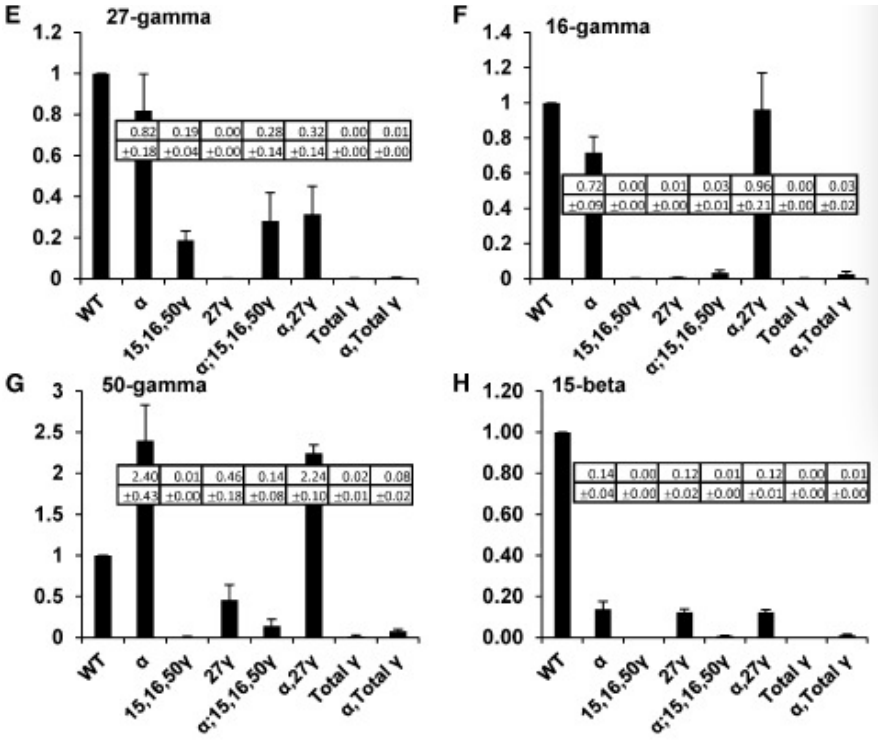


Figure 2. qRT-PCR expression analysis of α- and γ-zein genes in RNAi lines. (Source: Guo et al., 2013) Each graph represents the measurement of a different zein gene in all the RNAi lines relative to wild-type non-transgenic (HiII), which are shown on X axes. Values and SD (shown in insert) are the average of three biological replicate kernels and also incorporated three technical replicates of each measurement.

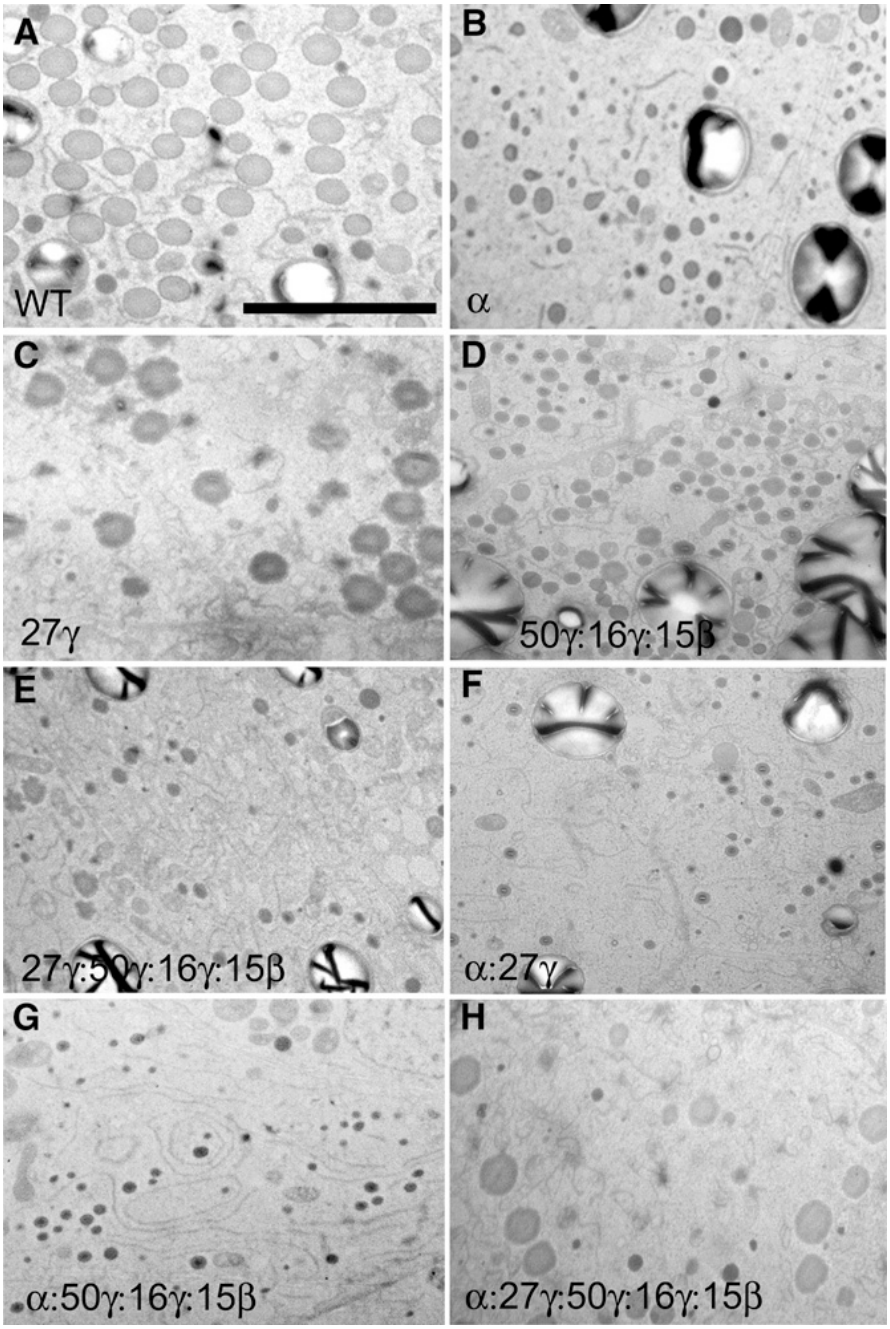


Figure. 3 TEM analysis showing protein body density in fourth sub-aleurone starchy layer of 18-DAP endosperm in zein RNAi lines and their crosses. (Source: Guo et al., 2013) Bar = 5 mm (A; refers to all panels). RNAi transgenes present are shown in bottom left of each section.

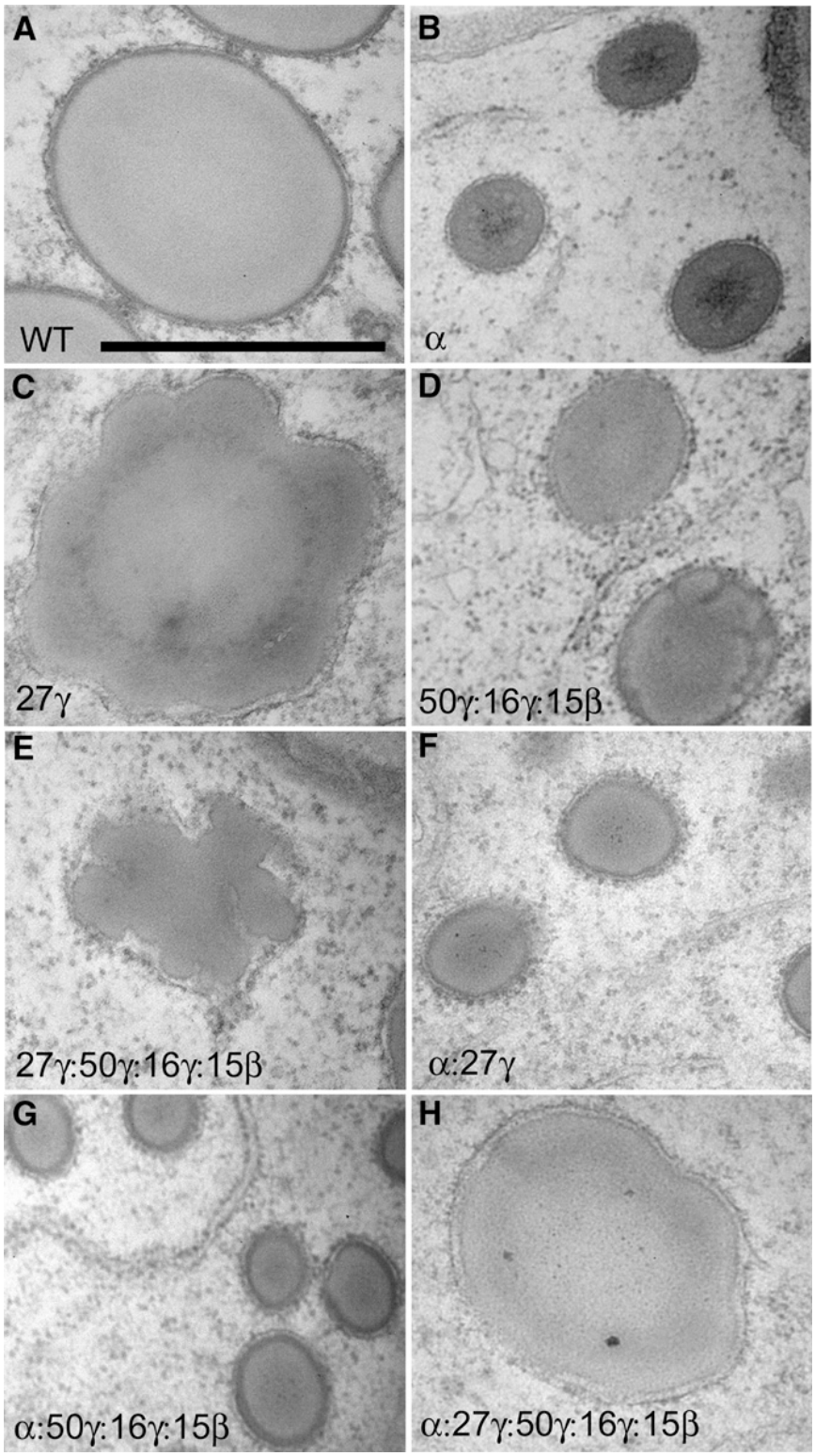


Figure. 4 TEM analysis showing protein size and morphology in fourth sub-aleurone starchy layer of 18-DAP endosperm in zein RNAi lines and their crosses. (Source: Guo et al., 2013) Bar = 1 mm (A; refers to all panels). RNAi transgenes present are shown in bottom left of each section.

REFERENCES

Guo X, Yuan L, Chen H, Sato SJ, Clemente TE, Holding DR (2013) Nonredundant function of zeins and their correct stoichiometric ratio drive protein body formation in maize endosperm. Plant Physiol 162: 1359-1369

Wu Y, Messing J (2010) RNA interference-mediated change in protein body morphology and seed opacity through loss of different zein proteins. Plant Physiol 153: 337-347

APPENDIX 2

Characterization of a sorghum high digestibility high lysine opaque mutant

1 Prolamins in sorghum

Similar to maize, sorghum has poor protein quality because of the lack of essential amino acids such as lysine and tryptophan (Bressani and Rios, 1962; MacLeon Jr et al., 1981; Mertz et al., 1984). Compounding this problem, sorghum proteins have poor digestibility. These properties result in severe malnutrition in populations where sorghum is consumed as the primary protein source (MacLeon Jr et al., 1981). The predominant seed proteins in sorghum are alcohol soluble prolamins called kafirins. Kafirins comprise more than 70% of the total seed protein and serve as a repository of nitrogen for recycling into the proteins of the germinating seedling. Based on their molecular weight and sequence, they are classified into four groups, namely, α, β, γ, and δ. α-kafirins are the main sorghum protein and comprises about 80% of total kafirins in the sorghum endosperm (Esen, 1987; Shull et al., 1991). Similar to the maize zeins, kafirins are stored in the endoplasmic reticulum (ER) in spherical protein bodies (PBs), in which γ- and β-kafirins are located in the peripheral PB region and α- and δ-kafirins are encapsulated in the inner region (Shull et al., 1992). The γ- and β-kafirins are rich in cysteine that through interchain disulphide bonding, form a cross-linked shell that blocks the accessibility of the internal α-kafirins to hydrolytic enzymes (Hamaker et al., 1987; Oria et al., 2000; Duodu et al., 2003; Wong et al., 2009). Because of the high abundance of α-kafirins, the indigestibility and lack of essential amino acids reduces their nutritive value (Hamaker et al., 1987; Hamaker et al., 1994).

The sorghum mutant *P721Q* was induced by EMS mutagenesis of a wild type line, P721N (N for normal) and was described first in Mohan's PhD thesis in 1975. It has increased lysine concentration as a result of reduced kafirin accumulation (increased in lysine-rich non-kafirins). It also has increased protein digestibility and was thus called the *high digestibility high lysine* (*hdhl*) variant. This mutant provides significant potential to improve the utility of sorghum as a human staple and as a livestock feed, and has been shown to have market potential. The microstructure of endosperm protein bodies of *hdhl* were found to have deep invaginations or folds that in many cases reached the central area of the protein body, forming irregularly shaped lobes (Oria et al., 2000). It is believed that the altered structure of the protein bodies results in substantially higher protein digestibility because of the increased surface area and improved protease accessibility to the major storage protein α-kafirin. Even though the *hdhl* mutant had been generally studied and its potential demonstrated, the molecular nature of the mutation in *hdhl* was unknown until we characterized it. Building on knowledge from maize mutants, we took a directed approach and identified this high lysine high-digestible sorghum (Wu et al., 2013).

We used polymorphic SSR markers to follow the locus of 22-kD α-kafirin genes that were cloned and characterized previously (Xu et al., 2012). Indeed, the polymorphic allele from *hdhl* was tightly linked to the floury phenotype in the F1 and BC1 populations. Because kafirin genes are the result of recent tandem amplification in the genome and are thus highly similar, random genomic and complementary DNA (cDNA) sequencing of

22-kD α-kafirin genes and mRNAs in *hdhl* was conducted. A single kafirin gene with a missense mutation in the signal peptide was identified in the mutant. To confirm that this mutation renders the protein resistant to processing, we compared normal and mutant kafirin expression in transgenic maize. The mutant kafirin not only remained uncleaved 24-kD α-kafirin precursor in maize, but also triggered the unfolded protein response (UPR) and the formation of irregular PBs as observed in sorghum *hdhl* endosperm. Therefore, these results validated the direct approach to explain the mechanism of a valuable trait in agriculture and offer a molecular marker for introgressing such a trait into local germplasm, which may ultimately increase the acreage of sorghum with improved nutritional value.

2 Three amino acids are highly conserved in α-prolamin signal peptides of *Panicoideae*

The α-prolamins, including α-kafirins and α-zeins, arose late in the evolution of the prolamins (Xu et al., 2012). Signal peptides from α-prolamins of *Panicoideae* such as maize, coix, sorghum, sugarcane and pearl millet are 21 amino acids long, and are cleaved when the proteins are transported into the lumen of the ER for storage. When the signal peptide sequences were aligned, only three amino acids were completely conserved except for the start codon: leucine (Leu) at positions 11 and 14 and Ala at 21 (Figure. 1), which suggested that these three amino acids were the most critical to the function of the α-prolamin signal peptide. In the *fl2* mutant of maize, the Ala 21 is replaced with valine (Val) in a 22-kD α-zein (Figure. 2), resulting in the retention of the signal peptide at maturity, since the Ala at position 21 is critical for cleavage. As a

consequence, the immature zein is tethered to the PB membrane, resulting in irregular PBs.

3 *hdhl* has a missense mutation at the conserved 21st amino acid

It was previously determined that all 20 α-kafirin genes were clustered on Chr. 5 of the BTx623 genome (Xu and Messing, 2008). The BAC SB40L16 sequence comprises a cluster of 10 active 22-kD α-kafirin genes of BTx623 within 35 Kb. SSR markers were designed based on the BAC SB40L16 sequence and this segment, as a whole, was integrated into the maize genome following transformation. These transgenic maize seeds expressed and processed kafirins properly, and served as a positive control in our experiments. The selected SSR markers showed that the tandem kafirin cluster on BAC SB40L16 was tightly linked to seed floury phenotype and the recombination frequency was less than 1%. Universal primers based on the conserved 5'-UTR and the 3'-end of the coding sequence (CDS) regions of the 22-kD α-kafirin gene cluster was used to amplify, from genomic DNA, all copies from P721N and *hdhl* for DNA sequencing. Seven of 77 sequences had a single nucleotide G-A substitution at position 61, whereas no such mutation was found in 96 sequences from P721N. Because it was previously indicated that all 10 kafirin genes in the cluster are expressed (Song et al., 2004), these results are consistent with one out of ten gene copies of the kafirin cluster producing the mutant protein. We also analyzed cDNA sequences of the 22-kD α-kafirin genes from developing endosperms and confirmed that the mutant transcript was only present in *hdhl* (five out of 96). This missense mutation causes the last amino acid alanine (Ala) in the signal peptide to be replaced with threonine (Thr) (Figure. 3). However, the Ala at

position 21 is critical for cleavage, as it is conserved in all 19- and 22-kD α-prolamins in foxtail millet, sorghum, coix and maize (Figure. 1).

4 The *hdhl* mutation renders the α-kafirin resistant to signal peptide processing

The mutant sorghum α-kafirin transcript with missense mutation under the control of the maize 22-kD α-zein promoter was transformed into maize. The heterologous system excluded any other sorghum protein and sequence and demonstrated that this mutation indeed rendered the kafirin resistant to signal peptide processing. Western blot analysis with a 22-kD α-zein antibody showed that transgenic maize expressing the mutant kafirin gene had a protein band at 24-kD, above the processed 22-kD α-zein. Also we expressed the normal kafirin cluster in transgenic maize as a control (Song et al., 2004). This 24-kD band is absent in the control with the normal kafirin genes (Figure. 4).

Sorghum has immense value as a staple food item for humans in Africa and Asia. It can substitute for maize in areas with low water supply and serve those geographical areas with food. Sorghum also has high nitrogen utilization efficiency, thereby cutting expenses in fertilizer. However, the sorghum protein is poorly digested and lacks essential amino acids. Therefore, a lysine-enhanced and highly digestible sorghum variety *hdhl* has gained broad attention. We identified the molecular nature of the sorghum mutant. Given the variability of the phenotype and with the introduction of modifiers for kernel hardness, breeders cannot rely on phenotype for the introgression of combined traits into local germplasms. Using a single molecular marker to follow the point mutation in the signal peptide of the mutant kafirin will allow breeders to select

their material without relying on a phenotype.

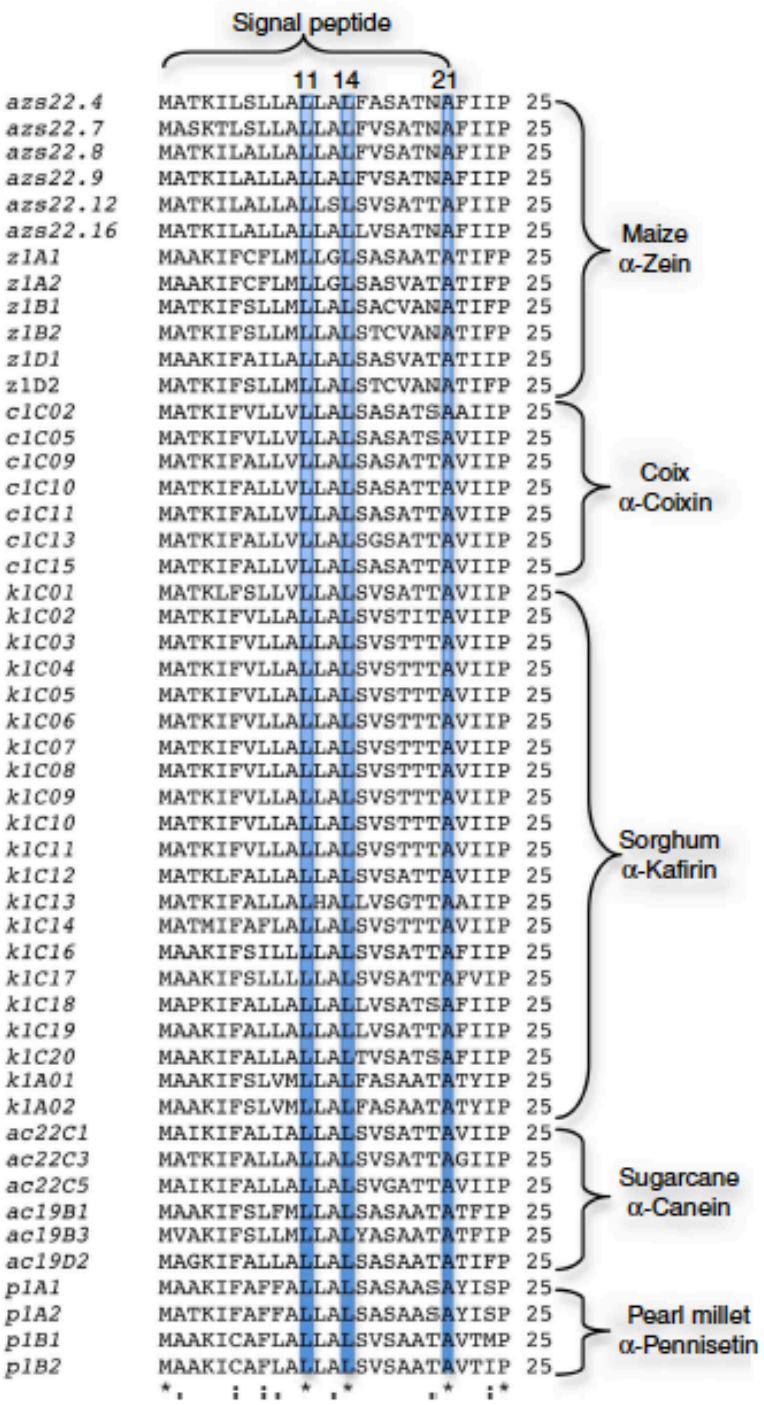


Figure. 1 Alignment of the signal peptides of α-prolamins from the grass subfamily. (Source: Wu et al., 2013)

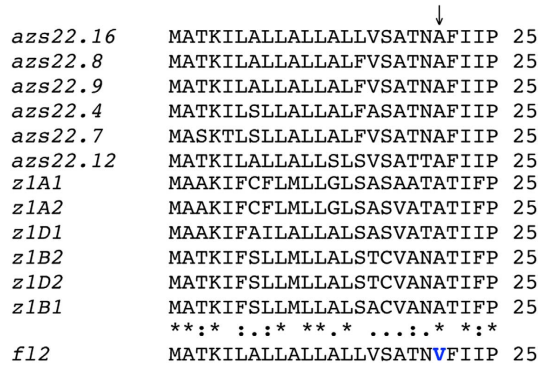


Figure. 2 Missense mutation at the 21st amino acid of the signal peptide in maize *floury2* mutants. (Source: Wu et al., 2013)

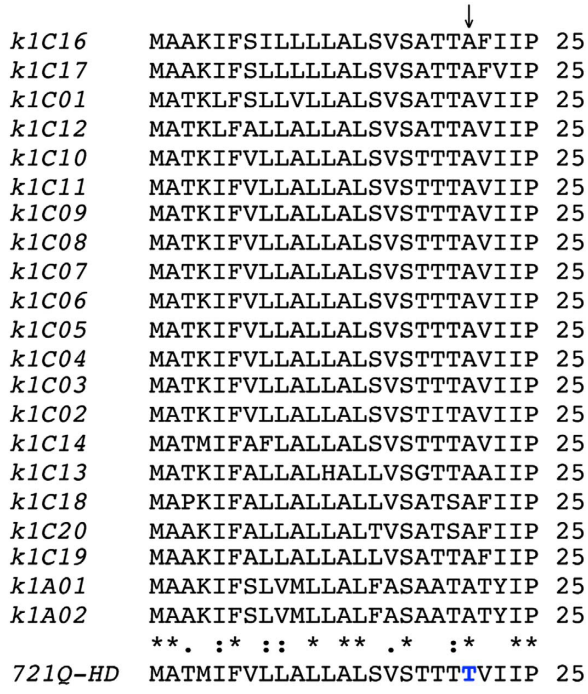


Figure. 3 Missense mutation at the 21st amino acid of the signal peptide in *hdhl*.
The missense mutation is marked with an arrow at the 21st amino acid of the signal peptide.
(Source: Wu et al., 2013)

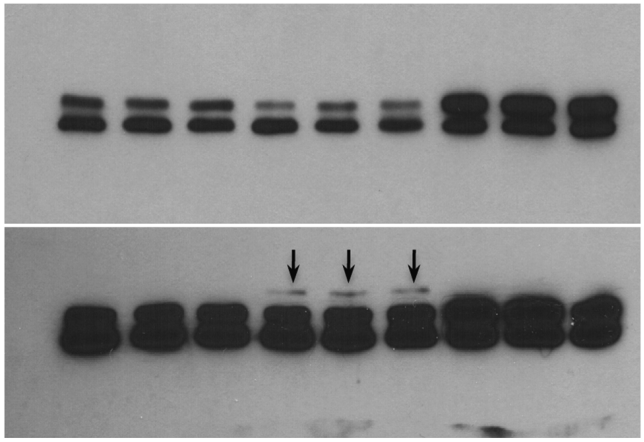


Figure. 4 Western blot analysis of kafirin accumulation. (Source: Wu et al., 2013). The antiserum can crossreact with the 22-kD zeins/kafirins and 19-kD α-zeins. The unprocessed mutant 22-kD α-kafirin (24 kD) migrated slower than the 22-kD α-zeins and normal 22-kD α-kafirins, as indicated by arrows. Upper panel, the film exposed for 10 s; bottom panel, the film exposed for 75 s. Lanes 1–3, null segregants; lanes 4–6, mutant kafirin transgenic seeds; lanes 7–9, transgenic seeds expressing normal 22-kD α-kafirins.

REFERENCES

Bressani R, Rios BJ (1962) The chemical and essential amino acid composition of twenty-five selections of grain sorghum. Cereal Chem 39: 50-58

Duodu K, Taylor J, Belton P, Hamaker B (2003) Factors affecting sorghum protein digestibility. Journal of Cereal Science 38: 117-131

Esen A (1987) A proposed nomenclature for the alcohol-soluble proteins (zeins) of maize. Journal of Cereal Science 5: 117-128

Hamaker B, Kirleis A, Butler L, Axtell J, Mertz E (1987) Improving the in vitro protein digestibility of sorghum with reducing agents. Proceedings of the National Academy of Sciences 84: 626-628

Hamaker B, Mertz E, Axtell J (1994) Effect of extrusion on sorghum kafirin solubility. Cereal chemistry 71: 515-515

Kim CS, Hunter BG, Kraft J, Boston RS, Yans S, Jung R, Larkins BA (2004) A defective signal peptide in a 19-kD alpha-zein protein causes the unfolded protein response and an opaque endosperm phenotype in the maize De*-B30 mutant. Plant Physiol 134: 380-387

Lending C, Larkins B (1992) Effect of the floury-2 locus on protein body formation during maize endosperm development. Protoplasma 171: 123-133

MacLeon Jr W, Lopez dRG, Placko RP, Graham GG (1981) Protein quality and digestibility of sorghum in preschool children: balance studies and plasma free amino acids. The Journal of nutrition 111: 1928-1936

Mertz ET, Hassen MM, Cairns-Whittern C, Kirleis AW, Tu L, Axtell JD (1984) Pepsin digestibility of proteins in sorghum and other major cereals. Proceedings of the National Academy of Sciences 81: 1-2

Oria MP, Hamaker BR, Axtell JD, Huang CP (2000) A highly digestible sorghum mutant cultivar exhibits a unique folded structure of endosperm protein bodies. Proc Natl Acad Sci U S A 97: 5065-5070

Shull JM, Watterson JJ, Kirleis AW (1991) Proposed nomenclature for the alcohol-soluble proteins (kafirins) of Sorghum bicolor (L. Moench) based on molecular weight, solubility, and structure. Journal of Agricultural and Food Chemistry 39: 83-87

Shull JM, Watterson JJ, Kirleis A (1992) Purification and immunocytochemical localization of kafirins inSorghum bicolor (L. Moench) endosperm. Protoplasma 171: 64-74

Song R, Segal G, Messing J (2004) Expression of the sorghum 10-member kafirin gene cluster in maize endosperm. Nucleic acids research 32: e189-e189

Wong JH, Lau T, Cai N, Singh J, Pedersen JF, Vensel WH, Hurkman WJ, Wilson JD, Lemaux PG, Buchanan BB (2009) Digestibility of protein and starch from sorghum is linked to biochemical and structural features of grain endosperm. Journal of cereal science 49: 73-82

Wu Y, Yuan L, Guo X, Holding DR, Messing J (2013) Mutation in the seed storage protein kafirin creates a high-value food trait in sorghum. Nature communications 4

Song R, Segal G, Messing J (2004) Expression of the sorghum 10-member kafirin gene cluster in maize endosperm. Nucleic acids research **32:** e189-e189

Xu J-H, Bennetzen JL, Messing J (2012) Dynamic gene copy number variation in collinear regions of grass genomes. Molecular biology and evolution 29: 861-871

Xu J-H, Messing J (2008) Organization of the prolamin gene family provides insight into the evolution of the maize genome and gene duplications in grass species. Proc Natl Acad Sci U S A 105: 14330-14335